Network Working Group
Request for Comments: 4907
Category: Informational

B. Aboba, Ed.
Internet Architecture Board
IAB

June 2007

Architectural Implications of Link Indications

Status of This Memo

Copyright Notice

Abstract

A link indication represents information provided by the link layer to higher layers regarding the state of the link. This document describes the role of link indications within the Internet architecture. While the judicious use of link indications can provide performance benefits, inappropriate use can degrade both robustness and performance. This document summarizes current proposals, describes the architectural issues, and provides examples of appropriate and inappropriate uses of link indications.

Table of Contents

1. Introduction

A link indication represents information provided by the link layer to higher layers regarding the state of the link. While the judicious use of link indications can provide performance benefits, inappropriate use can degrade both robustness and performance.

This document summarizes the current understanding of the role of link indications within the Internet architecture, and provides advice to document authors about the appropriate use of link indications within the Internet, transport, and application layers.

Section 1 describes the history of link indication usage within the Internet architecture and provides a model for the utilization of link indications. Section 2 describes the architectural considerations and provides advice to document authors. Section 3 describes recommendations and future work. Appendix A summarizes the literature on link indications, focusing largely on wireless Local Area Networks (WLANs).

1.1. Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

1.2. Terminology

Access Point (AP)
: A station that provides access to the fixed network (e.g., an 802.11 Distribution System), via the wireless medium (WM) for associated stations.

Asymmetric
: A link with transmission characteristics that are different depending upon the relative position or design characteristics of the transmitter and the receiver is said to be asymmetric. For instance, the range of one transmitter may be much higher than the range of another transmitter on the same medium.

Beacon
: A control message broadcast by a station (typically an Access Point), informing stations in the neighborhood of its continuing presence, possibly along with additional status or configuration information.

Binding Update (BU)
: A message indicating a mobile node's current mobility binding, and in particular its Care-of Address.

Correspondent Node
: A peer node with which a mobile node is communicating. The correspondent node may be either mobile or stationary.

Link
: A communication facility or medium over which nodes can communicate at the link layer, i.e., the layer immediately below the Internet Protocol (IP).

Link Down
: An event provided by the link layer that signifies a state change associated with the interface no longer being capable of communicating data frames; transient periods of high frame loss are not sufficient.

Link Indication
: Information provided by the link layer to higher layers regarding the state of the link.

Link Layer
: Conceptual layer of control or processing logic that is responsible for maintaining control of the link. The link layer functions provide an interface between the higher-layer logic and the link. The link layer is the layer immediately below the Internet Protocol (IP).

Link Up
: An event provided by the link layer that signifies a state change associated with the interface becoming capable of communicating data frames.

Maximum Segment Size (MSS)
: The maximum payload size available to the transport layer.

Maximum Transmission Unit (MTU)
: The size in octets of the largest IP packet, including the IP header and payload, that can be transmitted on a link or path.

Mobile Node
: A node that can change its point of attachment from one link to another, while still being reachable via its home address.

Operable Address
: A static or dynamically assigned address that has not been relinquished and has not expired.

Point of Attachment
: The endpoint on the link to which the host is currently connected.

Routable Address
: Any IP address for which routers will forward packets. This includes private addresses as specified in "Address Allocation for Private Internets" [RFC1918].

Station (STA)
: Any device that contains an IEEE 802.11 conformant medium access control (MAC) and physical layer (PHY) interface to the wireless medium (WM).

Strong End System Model
: The Strong End System model emphasizes the host/router distinction, tending to model a multi-homed host as a set of logical hosts within the same physical host. In the Strong End System model, addresses refer to an interface, rather than to the host to which they attach. As a result, packets sent on an outgoing interface have a source address configured on that interface, and incoming packets whose destination address does not correspond to the physical interface through which it is received are silently discarded.

Weak End System Model
: In the Weak End System model, addresses refer to a host. As a result, packets sent on an outgoing interface need not necessarily have a source address configured on that interface, and incoming packets whose destination address does not correspond to the physical interface through which it is received are accepted.

1.3. Overview

The use of link indications within the Internet architecture has a long history. In response to an attempt to send to a host that was off-line, the ARPANET link layer protocol provided a "Destination Dead" indication, described in "Fault Isolation and Recovery" [RFC816]. The ARPANET packet radio experiment [PRNET] incorporated frame loss in the calculation of routing metrics, a precursor to more recent link-aware routing metrics such as Expected Transmission Count (ETX), described in "A High-Throughput Path Metric for Multi-Hop Wireless Routing" [ETX].

"Routing Information Protocol" [RFC1058] defined RIP, which is descended from the Xerox Network Systems (XNS) Routing Information Protocol. "The OSPF Specification" [RFC1131] defined Open Shortest Path First, which uses Link State Advertisements (LSAs) in order to flood information relating to link status within an OSPF area. [RFC2328] defines version 2 of OSPF. While these and other routing protocols can utilize "Link Up" and "Link Down" indications provided by those links that support them, they also can detect link loss based on loss of routing packets. As noted in "Requirements for IP Version 4 Routers" [RFC1812]:

It is crucial that routers have workable mechanisms for determining that their network connections are functioning properly. Failure to detect link loss, or failure to take the proper actions when a problem is detected, can lead to black holes.

Attempts have also been made to define link indications other than "Link Up" and "Link Down". "Dynamically Switched Link Control Protocol" [RFC1307] defines an experimental protocol for control of links, incorporating "Down", "Coming Up", "Up", "Going Down", "Bring Down", and "Bring Up" states.

"A Generalized Model for Link Layer Triggers" [GenTrig] defines "generic triggers", including "Link Up", "Link Down", "Link Going Down", "Link Going Up", "Link Quality Crosses Threshold", "Trigger Rollback", and "Better Signal Quality AP Available". IEEE 802.21 [IEEE-802.21] defines a Media Independent Handover Event Service (MIH-ES) that provides event reporting relating to link characteristics, link status, and link quality. Events defined include "Link Down", "Link Up", "Link Going Down", "Link Signal Strength", and "Link Signal/Noise Ratio".

Under ideal conditions, links in the "up" state experience low frame loss in both directions and are immediately ready to send and receive data frames; links in the "down" state are unsuitable for sending and receiving data frames in either direction.

Unfortunately, links frequently exhibit non-ideal behavior. Wired links may fail in half-duplex mode, or exhibit partial impairment resulting in intermediate loss rates. Wireless links may exhibit asymmetry, intermittent frame loss, or rapid changes in throughput due to interference or signal fading. In both wired and wireless links, the link state may rapidly flap between the "up" and "down" states. This real-world behavior presents challenges to the integration of link indications with the Internet, transport, and application layers.

1.4. Layered Indication Model

A layered indication model is shown in Figure 1 that includes both internally generated link indications (such as link state and rate) and indications arising from external interactions such as path change detection. In this model, it is assumed that the link layer provides indications to higher layers primarily in the form of abstract indications that are link-technology agnostic.

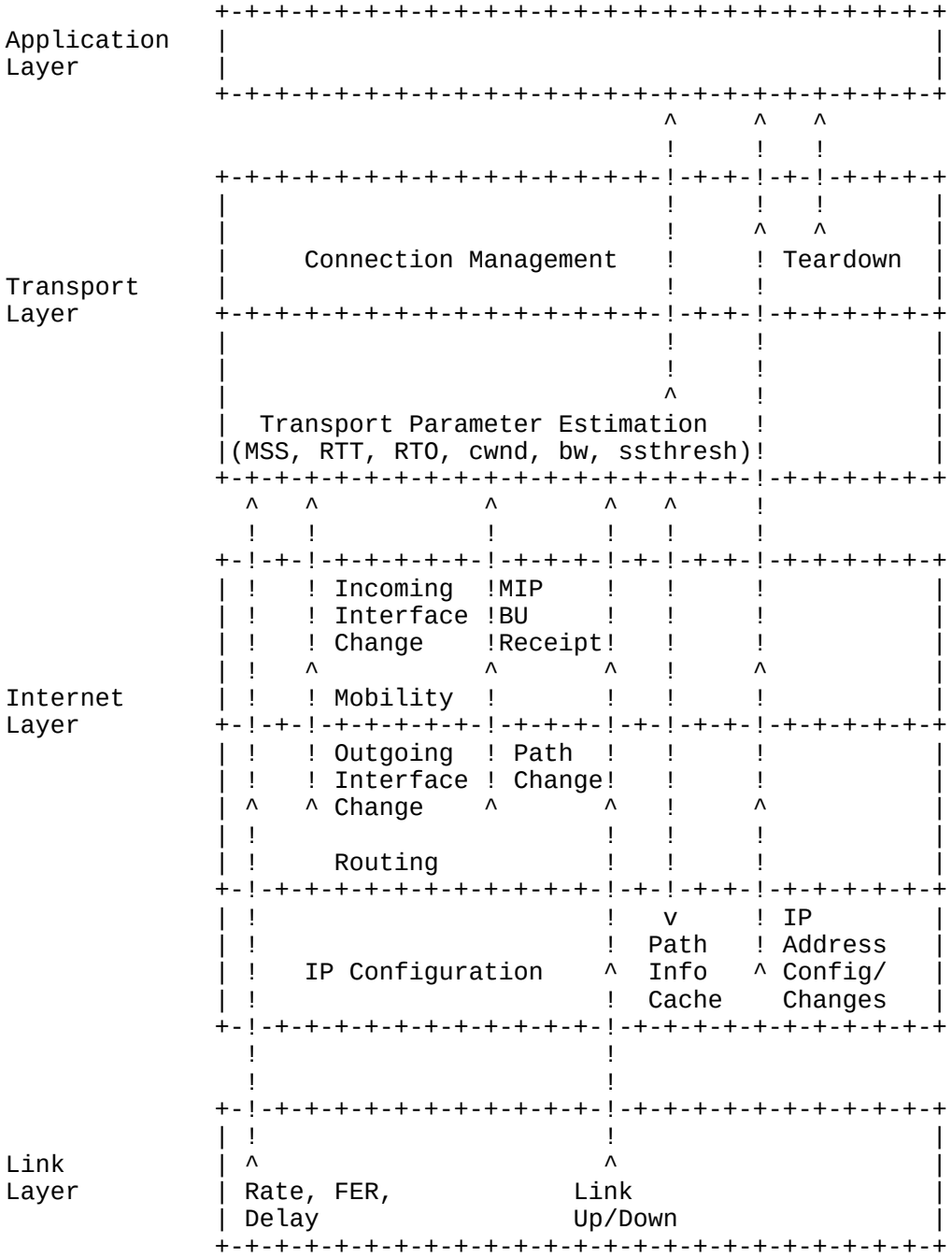

Figure 1. Layered Indication Model

1.4.1. Internet Layer

One of the functions of the Internet layer is to shield higher layers from the specifics of link behavior. As a result, the Internet layer validates and filters link indications and selects outgoing and incoming interfaces based on routing metrics.

The Internet layer composes its routing table based on information available from local interfaces as well as potentially by taking into account information provided by routers. This enables the state of the local routing table to reflect link conditions on both local and remote links. For example, prefixes to be added or removed from the routing table may be determined from Dynamic Host Configuration Protocol (DHCP) [RFC2131][RFC3315], Router Advertisements [RFC1256][RFC2461], redirect messages, or route updates incorporating information on the state of links multiple hops away.

As described in "Packetization Layer Path MTU Discovery" [RFC4821], the Internet layer may maintain a path information cache, enabling sharing of Path MTU information between concurrent or subsequent connections. The shared cache is accessed and updated by packetization protocols implementing packetization layer Path MTU Discovery.

The Internet layer also utilizes link indications in order to optimize aspects of Internet Protocol (IP) configuration and mobility. After receipt of a "Link Up" indication, hosts validate potential IP configurations by Detecting Network Attachment (DNA) [RFC4436]. Once the IP configuration is confirmed, it may be determined that an address change has occurred. However, "Link Up" indications may not necessarily result in a change to Internet layer configuration.

In "Detecting Network Attachment in IPv4" [RFC4436], after receipt of a "Link Up" indication, potential IP configurations are validated using a bidirectional reachability test. In "Detecting Network Attachment in IPv6 Networks (DNAv6)" [DNAv6], IP configuration is validated using reachability detection and Router Solicitation/Advertisement.

The routing sub-layer may utilize link indications in order to enable more rapid response to changes in link state and effective throughput. Link rate is often used in computing routing metrics. However, in wired networks the transmission rate may be negotiated in order to enhance energy efficiency [EfficientEthernet]. In wireless networks, the negotiated rate and Frame Error Rate (FER) may change

with link conditions so that effective throughput may vary on a packet-by-packet basis. In such situations, routing metrics may also exhibit rapid variation.

Routing metrics incorporating link indications such as Link Up/Down and effective throughput enable routers to take link conditions into account for the purposes of route selection. If a link experiences decreased rate or high frame loss, the route metric will increase for the prefixes that it serves, encouraging use of alternate paths if available. When the link condition improves, the route metric will decrease, encouraging use of the link.

Within Weak End System implementations, changes in routing metrics and link state may result in a change in the outgoing interface for one or more transport connections. Routes may also be added or withdrawn, resulting in loss or gain of peer connectivity. However, link indications such as changes in transmission rate or frame loss do not necessarily result in a change of outgoing interface.

The Internet layer may also become aware of path changes by other mechanisms, such as receipt of updates from a routing protocol, receipt of a Router Advertisement, dead gateway detection [RFC816] or network unreachability detection [RFC2461], ICMP redirects, or a change in the IPv4 TTL (Time to Live)/IPv6 Hop Limit of received packets. A change in the outgoing interface may in turn influence the mobility sub-layer, causing a change in the incoming interface. The mobility sub-layer may also become aware of a change in the incoming interface of a peer (via receipt of a Mobile IP Binding Update [RFC3775]).

1.4.2. Transport Layer

The transport layer processes received link indications differently for the purposes of transport parameter estimation and connection management.

For the purposes of parameter estimation, the transport layer is primarily interested in path properties that impact performance, and where link indications may be determined to be relevant to path properties they may be utilized directly. Link indications such as "Link Up"/"Link Down" or changes in rate, delay, and frame loss may prove relevant. This will not always be the case, however; where the bandwidth of the bottleneck on the end-to-end path is already much lower than the transmission rate, an increase in transmission rate may not materially affect path properties. As described in Appendix A.3, the algorithms for utilizing link layer indications to improve transport parameter estimates are still under development.

Strict layering considerations do not apply in transport path parameter estimation in order to enable the transport layer to make use of all available information. For example, the transport layer may determine that a link indication came from a link forming part of a path of one or more connections. In this case, it may utilize the receipt of a "Link Down" indication followed by a subsequent "Link Up" indication to infer the possibility of non-congestive packet loss during the period between the indications, even if the IP configuration does not change as a result, so that no Internet layer indication would be sent.

The transport layer may also find Internet layer indications useful for path parameter estimation. For example, path change indications can be used as a signal to reset path parameter estimates. Where there is no default route, loss of segments sent to a destination lacking a prefix in the local routing table may be assumed to be due to causes other than congestion, regardless of the reason for the removal (either because local link conditions caused it to be removed or because the route was withdrawn by a remote router).

For the purposes of connection management, layering considerations are important. The transport layer may tear down a connection based on Internet layer indications (such as a endpoint address changes), but does not take link indications into account. Just as a "Link Up" event may not result in a configuration change, and a configuration change may not result in connection teardown, the transport layer does not tear down connections on receipt of a "Link Down" indication, regardless of the cause. Where the "Link Down" indication results from frame loss rather than an explicit exchange, the indication may be transient, to be soon followed by a "Link Up" indication.

Even where the "Link Down" indication results from an explicit exchange such as receipt of a Point-to-Point Protocol (PPP) Link Control Protocol (LCP)-Terminate or an IEEE 802.11 Disassociate or Deauthenticate frame, an alternative point of attachment may be available, allowing connectivity to be quickly restored. As a result, robustness is best achieved by allowing connections to remain up until an endpoint address changes, or the connection is torn down due to lack of response to repeated retransmission attempts.

For the purposes of connection management, the transport layer is cautious with the use of Internet layer indications. Changes in the routing table are not relevant for the purposes of connection management, since it is desirable for connections to remain up during transitory routing flaps. However, the transport layer may tear down transport connections due to invalidation of a connection endpoint IP address. Where the connection has been established based on a Mobile

IP home address, a change in the Care-of Address need not result in connection teardown, since the configuration change is masked by the mobility functionality within the Internet layer, and is therefore transparent to the transport layer.

"Requirements for Internet Hosts -- Communication Layers" [RFC1122], Section 2.4, requires Destination Unreachable, Source Quench, Echo Reply, Timestamp Reply, and Time Exceeded ICMP messages to be passed up to the transport layer. [RFC1122], Section 4.2.3.9, requires Transmission Control Protocol (TCP) to react to an Internet Control Message Protocol (ICMP) Source Quench by slowing transmission.

[RFC1122], Section 4.2.3.9, distinguishes between ICMP messages indicating soft error conditions, which must not cause TCP to abort a connection, and hard error conditions, which should cause an abort. ICMP messages indicating soft error conditions include Destination Unreachable codes 0 (Net), 1 (Host), and 5 (Source Route Failed), which may result from routing transients; Time Exceeded; and Parameter Problem. ICMP messages indicating hard error conditions include Destination Unreachable codes 2 (Protocol Unreachable), 3 (Port Unreachable), and 4 (Fragmentation Needed and Don't Fragment Was Set). Since hosts implementing classical ICMP-based Path MTU Discovery [RFC1191] use Destination Unreachable code 4, they do not treat this as a hard error condition. Hosts implementing "Path MTU Discovery for IP version 6" [RFC1981] utilize ICMPv6 Packet Too Big messages. As noted in "TCP Problems with Path MTU Discovery" [RFC2923], classical Path MTU Discovery is vulnerable to failure if ICMP messages are not delivered or processed. In order to address this problem, "Packetization Layer Path MTU Discovery" [RFC4821] does depend on the delivery of ICMP messages.

"Fault Isolation and Recovery" [RFC816], Section 6, states:

It is not obvious, when error messages such as ICMP Destination Unreachable arrive, whether TCP should abandon the connection. The reason that error messages are difficult to interpret is that, as discussed above, after a failure of a gateway or network, there is a transient period during which the gateways may have incorrect information, so that irrelevant or incorrect error messages may sometimes return. An isolated ICMP Destination Unreachable may arrive at a host, for example, if a packet is sent during the period when the gateways are trying to find a new route. To abandon a TCP connection based on such a message arriving would be to ignore the valuable feature of the Internet that for many internal failures it reconstructs its function without any disruption of the end points.

"Requirements for IP Version 4 Routers" [RFC1812], Section 4.3.3.3, states that "Research seems to suggest that Source Quench consumes network bandwidth but is an ineffective (and unfair) antidote to congestion", indicating that routers should not originate them. In general, since the transport layer is able to determine an appropriate (and conservative) response to congestion based on packet loss or explicit congestion notification, ICMP Source Quench indications are not needed, and the sending of additional Source Quench packets during periods of congestion may be detrimental.

"ICMP attacks against TCP" [Gont] argues that accepting ICMP messages based on a correct four-tuple without additional security checks is ill-advised. For example, an attacker forging an ICMP hard error message can cause one or more transport connections to abort. The authors discuss a number of precautions, including mechanisms for validating ICMP messages and ignoring or delaying response to hard error messages under various conditions. They also recommend that hosts ignore ICMP Source Quench messages.

The transport layer may also provide information to the link layer. For example, the transport layer may wish to control the maximum number of times that a link layer frame may be retransmitted, so that the link layer does not continue to retransmit after a transport layer timeout. In IEEE 802.11, this can be achieved by adjusting the Management Information Base (MIB) [IEEE-802.11] variables dot11ShortRetryLimit (default: 7) and dot11LongRetryLimit (default: 4), which control the maximum number of retries for frames shorter and longer in length than dot11RTSThreshold, respectively. However, since these variables control link behavior as a whole they cannot be used to separately adjust behavior on a per-transport connection basis. In situations where the link layer retransmission timeout is of the same order as the path round-trip timeout, link layer control may not be possible at all.

1.4.3. Application Layer

The transport layer provides indications to the application layer by propagating Internet layer indications (such as IP address configuration and changes), as well as providing its own indications, such as connection teardown.

Since applications can typically obtain the information they need more reliably from the Internet and transport layers, they will typically not need to utilize link indications. A "Link Up" indication implies that the link is capable of communicating IP packets, but does not indicate that it has been configured; applications should use an Internet layer "IP Address Configured" event instead. "Link Down" indications are typically not useful to

applications, since they can be rapidly followed by a "Link Up" indication; applications should respond to transport layer teardown indications instead. Similarly, changes in the transmission rate may not be relevant to applications if the bottleneck bandwidth on the path does not change; the transport layer is best equipped to determine this. As a result, Figure 1 does not show link indications being provided directly to applications.

2. Architectural Considerations

The complexity of real-world link behavior poses a challenge to the integration of link indications within the Internet architecture. While the literature provides persuasive evidence of the utility of link indications, difficulties can arise in making effective use of them. To avoid these issues, the following architectural principles are suggested and discussed in more detail in the sections that follow:

(1) Proposals should avoid use of simplified link models in circumstances where they do not apply (Section 2.1).

(2) Link indications should be clearly defined, so that it is understood when they are generated on different link layers (Section 2.2).

(3) Proposals must demonstrate robustness against spurious link indications (Section 2.3).

(4) Upper layers should utilize a timely recovery step so as to limit the potential damage from link indications determined to be invalid after they have been acted on (Section 2.3.2).

(5) Proposals must demonstrate that effective congestion control is maintained (Section 2.4).

(6) Proposals must demonstrate the effectiveness of proposed optimizations (Section 2.5).

(7) Link indications should not be required by upper layers, in order to maintain link independence (Section 2.6).

(8) Proposals should avoid race conditions, which can occur where link indications are utilized directly by multiple layers of the stack (Section 2.7).

(9) Proposals should avoid inconsistencies between link and routing layer metrics (Section 2.7.3).

(10) Overhead reduction schemes must avoid compromising interoperability and introducing link layer dependencies into the Internet and transport layers (Section 2.8).

(11) Proposals for transport of link indications beyond the local host need to carefully consider the layering, security, and transport implications (Section 2.9).

2.1. Model Validation

Proposals should avoid the use of link models in circumstances where they do not apply.

In "The mistaken axioms of wireless-network research" [Kotz], the authors conclude that mistaken assumptions relating to link behavior may lead to the design of network protocols that may not work in practice. For example, the authors note that the three-dimensional nature of wireless propagation can result in large signal strength changes over short distances. This can result in rapid changes in link indications such as rate, frame loss, and signal strength.

In "Modeling Wireless Links for Transport Protocols" [GurtovFloyd], the authors provide examples of modeling mistakes and examples of how to improve modeling of link characteristics. To accompany the paper, the authors provide simulation scenarios in ns-2.

In order to avoid the pitfalls described in [Kotz] [GurtovFloyd], documents that describe capabilities that are dependent on link indications should explicitly articulate the assumptions of the link model and describe the circumstances in which they apply.

Generic "trigger" models may include implicit assumptions that may prove invalid in outdoor or mesh wireless LAN deployments. For example, two-state Markov models assume that the link is either in a state experiencing low frame loss ("up") or in a state where few frames are successfully delivered ("down"). In these models, symmetry is also typically assumed, so that the link is either "up" in both directions or "down" in both directions. In situations where intermediate loss rates are experienced, these assumptions may be invalid.

As noted in "Hybrid Rate Control for IEEE 802.11" [Haratcherev], signal strength data is noisy and sometimes inconsistent, so that it needs to be filtered in order to avoid erratic results. Given this, link indications based on raw signal strength data may be unreliable. In order to avoid problems, it is best to combine signal strength data with other techniques. For example, in developing a "Going Down" indication for use with [IEEE-802.21] it would be advisable to

validate filtered signal strength measurements with other indications of link loss such as lack of Beacon reception.

2.2. Clear Definitions

Link indications should be clearly defined, so that it is understood when they are generated on different link layers. For example, considerable work has been required in order to come up with the definitions of "Link Up" and "Link Down", and to define when these indications are sent on various link layers.

Link indication definitions should heed the following advice:

(1) Do not assume symmetric link performance or frame loss that is either low ("up") or high ("down").

In wired networks, links in the "up" state typically experience low frame loss in both directions and are ready to send and receive data frames; links in the "down" state are unsuitable for sending and receiving data frames in either direction. Therefore, a link providing a "Link Up" indication will typically experience low frame loss in both directions, and high frame loss in any direction can only be experienced after a link provides a "Link Down" indication. However, these assumptions may not hold true for wireless LAN networks. Asymmetry is typically less of a problem for cellular networks where propagation occurs over longer distances, multi-path effects may be less severe, and the base station can transmit at much higher power than mobile stations while utilizing a more sensitive antenna.

Specifications utilizing a "Link Up" indication should not assume that receipt of this indication means that the link is experiencing symmetric link conditions or low frame loss in either direction. In general, a "Link Up" event should not be sent due to transient changes in link conditions, but only due to a change in link layer state. It is best to assume that a "Link Up" event may not be sent in a timely way. Large handoff latencies can result in a delay in the generation of a "Link Up" event as movement to an alternative point of attachment is delayed.

(2) Consider the sensitivity of link indications to transient link conditions. Due to common effects such as multi-path interference, signal strength and signal to noise ratio (SNR) may vary rapidly over a short distance, causing erratic behavior of link indications based on unfiltered measurements. As noted in [Haratcherev], signal strength may prove most useful when

utilized in combination with other measurements, such as frame loss.

(3) Where possible, design link indications with built-in damping. By design, the "Link Up" and "Link Down" events relate to changes in the state of the link layer that make it able and unable to communicate IP packets. These changes are generated either by the link layer state machine based on link layer exchanges (e.g., completion of the IEEE 802.11i four-way handshake for "Link Up", or receipt of a PPP LCP-Terminate for "Link Down") or by protracted frame loss, so that the link layer concludes that the link is no longer usable. As a result, these link indications are typically less sensitive to changes in transient link conditions.

(4) Do not assume that a "Link Down" event will be sent at all, or that, if sent, it will be received in a timely way. A good link layer implementation will both rapidly detect connectivity failure (such as by tracking missing Beacons) while sending a "Link Down" event only when it concludes the link is unusable, not due to transient frame loss.

However, existing wireless LAN implementations often do not do a good job of detecting link failure. During a lengthy detection phase, a "Link Down" event is not sent by the link layer, yet IP packets cannot be transmitted or received on the link. Initiation of a scan may be delayed so that the station cannot find another point of attachment. This can result in inappropriate backoff of retransmission timers within the transport layer, among other problems. This is not as much of a problem for cellular networks that utilize transmit power adjustment.

2.3. Robustness

Link indication proposals must demonstrate robustness against misleading indications. Elements to consider include:

Implementation variation
Recovery from invalid indications
Damping and hysteresis

2.3.1. Implementation Variation

Variations in link layer implementations may have a substantial impact on the behavior of link indications. These variations need to be taken into account in evaluating the performance of proposals. For example, radio propagation and implementation differences can impact the reliability of link indications.

In "Link-level Measurements from an 802.11b Mesh Network" [Aguayo], the authors analyze the cause of frame loss in a 38-node urban multi-hop IEEE 802.11 ad-hoc network. In most cases, links that are very bad in one direction tend to be bad in both directions, and links that are very good in one direction tend to be good in both directions. However, 30 percent of links exhibited loss rates differing substantially in each direction.

As described in [Aguayo], wireless LAN links often exhibit loss rates intermediate between "up" (low loss) and "down" (high loss) states, as well as substantial asymmetry. As a result, receipt of a "Link Up" indication may not necessarily indicate bidirectional reachability, since it could have been generated after exchange of small frames at low rates, which might not imply bidirectional connectivity for large frames exchanged at higher rates.

Where multi-path interference or hidden nodes are encountered, signal strength may vary widely over a short distance. Several techniques may be used to reduce potential disruptions. Multiple transmitting and receiving antennas may be used to reduce multi-path effects; transmission rate adaptation can be used to find a more satisfactory transmission rate; transmit power adjustment can be used to improve signal quality and reduce interference; Request-to-Send/Clear-to-Send (RTS/CTS) signaling can be used to reduce hidden node problems. These techniques may not be completely effective, so that high frame loss may be encountered, causing the link to cycle between "up" and "down" states.

To improve robustness against spurious link indications, it is recommended that upper layers treat the indication as a "hint" (advisory in nature), rather than a "trigger" dictating a particular action. Upper layers may then attempt to validate the hint.

In [RFC4436], "Link Up" indications are rate limited, and IP configuration is confirmed using bidirectional reachability tests carried out coincident with a request for configuration via DHCP. As a result, bidirectional reachability is confirmed prior to activation of an IP configuration. However, where a link exhibits an intermediate loss rate, demonstration of bidirectional reachability may not necessarily indicate that the link is suitable for carrying IP data packets.

Another example of validation occurs in IPv4 Link-Local address configuration [RFC3927]. Prior to configuration of an IPv4 Link-Local address, it is necessary to run a claim-and-defend protocol. Since a host needs to be present to defend its address against another claimant, and address conflicts are relatively likely, a host returning from sleep mode or receiving a "Link Up" indication could

encounter an address conflict were it to utilize a formerly configured IPv4 Link-Local address without rerunning claim and defend.

2.3.2. Recovery from Invalid Indications

In some situations, improper use of link indications can result in operational malfunctions. It is recommended that upper layers utilize a timely recovery step so as to limit the potential damage from link indications determined to be invalid after they have been acted on.

In Detecting Network Attachment in IPv4 (DNAv4) [RFC4436], reachability tests are carried out coincident with a request for configuration via DHCP. Therefore, if the bidirectional reachability test times out, the host can still obtain an IP configuration via DHCP, and if that fails, the host can still continue to use an existing valid address if it has one.

Where a proposal involves recovery at the transport layer, the recovered transport parameters (such as the Maximum Segment Size (MSS), RoundTrip Time (RTT), Retransmission TimeOut (RTO), Bandwidth (bw), congestion window (cwnd), etc.) should be demonstrated to remain valid. Congestion window validation is discussed in "TCP Congestion Window Validation" [RFC2861].

Where timely recovery is not supported, unexpected consequences may result. As described in [RFC3927], early IPv4 Link-Local implementations would wait five minutes before attempting to obtain a routable address after assigning an IPv4 Link-Local address. In one implementation, it was observed that where mobile hosts changed their point of attachment more frequently than every five minutes, they would never obtain a routable address. The problem was caused by an invalid link indication (signaling of "Link Up" prior to completion of link layer authentication), resulting in an initial failure to obtain a routable address using DHCP. As a result, [RFC3927] recommends against modification of the maximum retransmission timeout (64 seconds) provided in [RFC2131].

2.3.3. Damping and Hysteresis

Damping and hysteresis can be utilized to limit damage from unstable link indications. This may include damping unstable indications or placing constraints on the frequency of link indication-induced actions within a time period.

While [Aguayo] found that frame loss was relatively stable for stationary stations, obstacles to radio propagation and multi-path interference can result in rapid changes in signal strength for a mobile station. As a result, it is possible for mobile stations to encounter rapid changes in link characteristics, including changes in transmission rate, throughput, frame loss, and even "Link Up"/"Link Down" indications.

Where link-aware routing metrics are implemented, this can result in rapid metric changes, potentially resulting in frequent changes in the outgoing interface for Weak End System implementations. As a result, it may be necessary to introduce route flap dampening.

However, the benefits of damping need to be weighed against the additional latency that can be introduced. For example, in order to filter out spurious "Link Down" indications, these indications may be delayed until it can be determined that a "Link Up" indication will not follow shortly thereafter. However, in situations where multiple Beacons are missed such a delay may not be needed, since there is no evidence of a suitable point of attachment in the vicinity.

In some cases, it is desirable to ignore link indications entirely. Since it is possible for a host to transition from an ad-hoc network to a network with centralized address management, a host receiving a "Link Up" indication cannot necessarily conclude that it is appropriate to configure an IPv4 Link-Local address prior to determining whether a DHCP server is available [RFC3927] or an operable configuration is valid [RFC4436].

As noted in Section 1.4, the transport layer does not utilize "Link Up" and "Link Down" indications for the purposes of connection management.

2.4. Congestion Control

Link indication proposals must demonstrate that effective congestion control is maintained [RFC2914]. One or more of the following techniques may be utilized:

- Rate limiting. Packets generated based on receipt of link indications can be rate limited (e.g., a limit of one packet per end-to-end path RTO).

- Utilization of upper-layer indications. Applications should depend on upper-layer indications such as IP address configuration/change notification, rather than utilizing link indications such as "Link Up".

Keepalives. In order to improve robustness against spurious link indications, an application keepalive or transport layer indication (such as connection teardown) can be used instead of consuming "Link Down" indications.

Conservation of resources. Proposals must demonstrate that they are not vulnerable to congestive collapse.

As noted in "Robust Rate Adaptation for 802.11 Wireless Networks" [Robust], decreasing transmission rate in response to frame loss increases contention, potentially leading to congestive collapse. To avoid this, the link layer needs to distinguish frame loss due to congestion from loss due to channel conditions. Only frame loss due to deterioration in channel conditions can be used as a basis for decreasing transmission rate.

Consider a proposal where a "Link Up" indication is used by a host to trigger retransmission of the last previously sent packet, in order to enable ACK reception prior to expiration of the host's retransmission timer. On a rapidly moving mobile node where "Link Up" indications follow in rapid succession, this could result in a burst of retransmitted packets, violating the principle of "conservation of packets".

At the application layer, link indications have been utilized by applications such as Presence [RFC2778] in order to optimize registration and user interface update operations. For example, implementations may attempt presence registration on receipt of a "Link Up" indication, and presence de-registration by a surrogate receiving a "Link Down" indication. Presence implementations using "Link Up"/"Link Down" indications this way violate the principle of "conservation of packets" since link indications can be generated on a time scale less than the end-to-end path RTO. The problem is magnified since for each presence update, notifications can be delivered to many watchers. In addition, use of a "Link Up" indication in this manner is unwise since the interface may not yet even have an operable Internet layer configuration. Instead, an "IP address configured" indication may be utilized.

2.5. Effectiveness

Proposals must demonstrate the effectiveness of proposed optimizations. Since optimizations typically increase complexity, substantial performance improvement is required in order to make a compelling case.

In the face of unreliable link indications, effectiveness may depend on the penalty for false positives and false negatives. In the case of DNAv4 [RFC4436], the benefits of successful optimization are modest, but the penalty for being unable to confirm an operable configuration is a lengthy timeout. As a result, the recommended strategy is to test multiple potential configurations in parallel in addition to attempting configuration via DHCP. This virtually guarantees that DNAv4 will always result in performance equal to or better than use of DHCP alone.

2.6. Interoperability

While link indications can be utilized where available, they should not be required by upper layers, in order to maintain link layer independence. For example, if information on supported prefixes is provided at the link layer, hosts not understanding those hints must still be able to obtain an IP address.

Where link indications are proposed to optimize Internet layer configuration, proposals must demonstrate that they do not compromise robustness by interfering with address assignment or routing protocol behavior, making address collisions more likely, or compromising Duplicate Address Detection (DAD) [RFC4429].

To avoid compromising interoperability in the pursuit of performance optimization, proposals must demonstrate that interoperability remains possible (potentially with degraded performance) even if one or more participants do not implement the proposal.

2.7. Race Conditions

Link indication proposals should avoid race conditions, which can occur where link indications are utilized directly by multiple layers of the stack.

Link indications are useful for optimization of Internet Protocol layer addressing and configuration as well as routing. Although "The BU-trigger method for improving TCP performance over Mobile IPv6" [Kim] describes situations in which link indications are first processed by the Internet Protocol layer (e.g., MIPv6) before being utilized by the transport layer, for the purposes of parameter estimation, it may be desirable for the transport layer to utilize link indications directly.

In situations where the Weak End System model is implemented, a change of outgoing interface may occur at the same time the transport layer is modifying transport parameters based on other link

indications. As a result, transport behavior may differ depending on the order in which the link indications are processed.

Where a multi-homed host experiences increasing frame loss or decreased rate on one of its interfaces, a routing metric taking these effects into account will increase, potentially causing a change in the outgoing interface for one or more transport connections. This may trigger Mobile IP signaling so as to cause a change in the incoming path as well. As a result, the transport parameters estimated for the original outgoing and incoming paths (congestion state, Maximum Segment Size (MSS) derived from the link maximum transmission unit (MTU) or Path MTU) may no longer be valid for the new outgoing and incoming paths.

To avoid race conditions, the following measures are recommended:

Path change re-estimation
Layering
Metric consistency

2.7.1. Path Change Re-estimation

When the Internet layer detects a path change, such as a major change in transmission rate, a change in the outgoing or incoming interface of the host or the incoming interface of a peer, or perhaps even a substantial change in the IPv4 TTL/IPv6 Hop Limit of received packets, it may be worth considering whether to reset transport parameters (RTT, RTO, cwnd, bw, MSS) to their initial values so as to allow them to be re-estimated. This ensures that estimates based on the former path do not persist after they have become invalid. Appendix A.3 summarizes the research on this topic.

2.7.2. Layering

Another technique to avoid race conditions is to rely on layering to damp transient link indications and provide greater link layer independence.

The Internet layer is responsible for routing as well as IP configuration and mobility, providing higher layers with an abstraction that is independent of link layer technologies.

In general, it is advisable for applications to utilize indications from the Internet or transport layers rather than consuming link indications directly.

2.7.3. Metric Consistency

Proposals should avoid inconsistencies between link and routing layer metrics. Without careful design, potential differences between link indications used in routing and those used in roaming and/or link enablement can result in instability, particularly in multi-homed hosts.

Once a link is in the "up" state, its effectiveness in transmission of data packets can be used to determine an appropriate routing metric. In situations where the transmission time represents a large portion of the total transit time, minimizing total transmission time is equivalent to maximizing effective throughput. "A High-Throughput Path Metric for Multi-Hop Wireless Routing" [ETX] describes a proposed routing metric based on the Expected Transmission Count (ETX). The authors demonstrate that ETX, based on link layer frame loss rates (prior to retransmission), enables the selection of routes maximizing effective throughput. Where the transmission rate is constant, the expected transmission time is proportional to ETX, so that minimizing ETX also minimizes expected transmission time.

However, where the transmission rate may vary, ETX may not represent a good estimate of the estimated transmission time. In "Routing in multi-radio, multi-hop wireless mesh networks" [ETX-Rate], the authors define a new metric called Expected Transmission Time (ETT). This is described as a "bandwidth adjusted ETX" since ETT = ETX * S/B where S is the size of the probe packet and B is the bandwidth of the link as measured by a packet pair [Morgan]. However, ETT assumes that the loss fraction of small probe frames sent at 1 Mbps data rate is indicative of the loss fraction of larger data frames at higher rates, which tends to underestimate the ETT at higher rates, where frame loss typically increases. In "A Radio Aware Routing Protocol for Wireless Mesh Networks" [ETX-Radio], the authors refine the ETT metric further by estimating the loss fraction as a function of transmission rate.

However, prior to sending data packets over the link, the appropriate routing metric may not easily be predicted. As noted in [Shortest], a link that can successfully transmit the short frames utilized for control, management, or routing may not necessarily be able to reliably transport larger data packets.

Therefore, it may be necessary to utilize alternative metrics (such as signal strength or Access Point load) in order to assist in attachment/handoff decisions. However, unless the new interface is the preferred route for one or more destination prefixes, a Weak End System implementation will not use the new interface for outgoing traffic. Where "idle timeout" functionality is implemented, the

unused interface will be brought down, only to be brought up again by the link enablement algorithm.

Within the link layer, metrics such as signal strength and frame loss may be used to determine the transmission rate, as well as to determine when to select an alternative point of attachment. In order to enable stations to roam prior to encountering packet loss, studies such as "An experimental study of IEEE 802.11b handover performance and its effect on voice traffic" [Vatn] have suggested using signal strength as a mechanism to more rapidly detect loss of connectivity, rather than frame loss, as suggested in "Techniques to Reduce IEEE 802.11b MAC Layer Handover Time" [Velayos].

[Aguayo] notes that signal strength and distance are not good predictors of frame loss or throughput, due to the potential effects of multi-path interference. As a result, a link brought up due to good signal strength may subsequently exhibit significant frame loss and a low throughput. Similarly, an Access Point (AP) demonstrating low utilization may not necessarily be the best choice, since utilization may be low due to hardware or software problems. "OSPF Optimized Multipath (OSPF-OMP)" [Villamizar] notes that link-utilization-based routing metrics have a history of instability.

2.8. Layer Compression

In many situations, the exchanges required for a host to complete a handoff and reestablish connectivity are considerable, leading to proposals to combine exchanges occurring within multiple layers in order to reduce overhead. While overhead reduction is a laudable goal, proposals need to avoid compromising interoperability and introducing link layer dependencies into the Internet and transport layers.

Exchanges required for handoff and connectivity reestablishment may include link layer scanning, authentication, and association establishment; Internet layer configuration, routing, and mobility exchanges; transport layer retransmission and recovery; security association reestablishment; application protocol re-authentication and re-registration exchanges, etc.

Several proposals involve combining exchanges within the link layer. For example, in [EAPIKEv2], a link layer Extensible Authentication Protocol (EAP) [RFC3748] exchange may be used for the purpose of IP address assignment, potentially bypassing Internet layer configuration. Within [PEAP], it is proposed that a link layer EAP exchange be used for the purpose of carrying Mobile IPv6 Binding Updates. [MIPEAP] proposes that EAP exchanges be used for configuration of Mobile IPv6. Where link, Internet, or transport

layer mechanisms are combined, hosts need to maintain backward compatibility to permit operation on networks where compression schemes are not available.

Layer compression schemes may also negatively impact robustness. For example, in order to optimize IP address assignment, it has been proposed that prefixes be advertised at the link layer, such as within the 802.11 Beacon and Probe Response frames. However, [IEEE-802.1X] enables the Virtual LAN Identifier (VLANID) to be assigned dynamically, so that prefix(es) advertised within the Beacon and/or Probe Response may not correspond to the prefix(es) configured by the Internet layer after the host completes link layer authentication. Were the host to handle IP configuration at the link layer rather than within the Internet layer, the host might be unable to communicate due to assignment of the wrong IP address.

2.9. Transport of Link Indications

Proposals for the transport of link indications need to carefully consider the layering, security, and transport implications.

As noted earlier, the transport layer may take the state of the local routing table into account in improving the quality of transport parameter estimates. While absence of positive feedback that the path is sending data end-to-end must be heeded, where a route that had previously been absent is recovered, this may be used to trigger congestion control probing. While this enables transported link indications that affect the local routing table to improve the quality of transport parameter estimates, security and interoperability considerations relating to routing protocols still apply.

Proposals involving transport of link indications need to demonstrate the following:

(a) Superiority to implicit signals. In general, implicit signals are preferred to explicit transport of link indications since they do not require participation in the routing mesh, add no new packets in times of network distress, operate more reliably in the presence of middle boxes such as NA(P)Ts, are more likely to be backward compatible, and are less likely to result in security vulnerabilities. As a result, explicit signaling proposals must prove that implicit signals are inadequate.

(b) Mitigation of security vulnerabilities. Transported link indications should not introduce new security vulnerabilities. Link indications that result in modifications to the local routing table represent a routing protocol, so that the

vulnerabilities associated with unsecured routing protocols apply, including spoofing by off-link attackers. While mechanisms such as "SEcure Neighbor Discovery (SEND)" [RFC3971] may enable authentication and integrity protection of router-originated messages, protecting against forgery of transported link indications, they are not yet widely deployed.

(c) Validation of transported indications. Even if a transported link indication can be integrity protected and authenticated, if the indication is sent by a host off the local link, it may not be clear that the sender is on the actual path in use, or which transport connection(s) the indication relates to. Proposals need to describe how the receiving host can validate the transported link indication.

(d) Mapping of Identifiers. When link indications are transported, it is generally for the purposes of providing information about Internet, transport, or application layer operations at a remote element. However, application layer sessions or transport connections may not be visible to the remote element due to factors such as load sharing between links, or use of IPsec, tunneling protocols, or nested headers. As a result, proposals need to demonstrate how the link indication can be mapped to the relevant higher-layer state. For example, on receipt of a link indication, the transport layer will need to identify the set of transport sessions (source address, destination address, source port, destination port, transport) that are affected. If a presence server is receiving remote indications of "Link Up"/"Link Down" status for a particular Media Access Control (MAC) address, the presence server will need to associate that MAC address with the identity of the user (pres:user@example.com) to whom that link status change is relevant.

3. Future Work

Further work is needed in order to understand how link indications can be utilized by the Internet, transport, and application layers.

More work is needed to understand the connection between link indications and routing metrics. For example, the introduction of block ACKs (supported in [IEEE-802.11e]) complicates the relationship between effective throughput and frame loss, which may necessitate the development of revised routing metrics for ad-hoc networks. More work is also needed to reconcile handoff metrics (e.g., signal strength and link utilization) with routing metrics based on link indications (e.g., frame error rate and negotiated rate).

A better understanding of the use of physical and link layer metrics in rate negotiation is required. For example, recent work [Robust][CARA] has suggested that frame loss due to contention (which would be exacerbated by rate reduction) can be distinguished from loss due to channel conditions (which may be improved via rate reduction).

At the transport layer, more work is needed to determine the appropriate reaction to Internet layer indications such as routing table and path changes. More work is also needed in utilization of link layer indications in transport parameter estimation, including rate changes, "Link Up"/"Link Down" indications, link layer retransmissions, and frame loss of various types (due to contention or channel conditions).

More work is also needed to determine how link layers may utilize information from the transport layer. For example, it is undesirable for a link layer to retransmit so aggressively that the link layer round-trip time approaches that of the end-to-end transport connection. Instead, it may make sense to do downward rate adjustment so as to decrease frame loss and improve latency. Also, in some cases, the transport layer may not require heroic efforts to avoid frame loss; timely delivery may be preferred instead.

4. Security Considerations

Proposals for the utilization of link indications may introduce new security vulnerabilities. These include:

- Spoofing
- Indication validation
- Denial of service

4.1. Spoofing

Where link layer control frames are unprotected, they may be spoofed by an attacker. For example, PPP does not protect LCP frames such as LCP-Terminate, and [IEEE-802.11] does not protect management frames such as Associate/Reassociate, Disassociate, or Deauthenticate.

Spoofing of link layer control traffic may enable attackers to exploit weaknesses in link indication proposals. For example, proposals that do not implement congestion avoidance can enable attackers to mount denial-of-service attacks.

However, even where the link layer incorporates security, attacks may still be possible if the security model is not consistent. For example, wireless LANs implementing [IEEE-802.11i] do not enable

stations to send or receive IP packets on the link until completion of an authenticated key exchange protocol known as the "4-way handshake". As a result, a link implementing [IEEE-802.11i] cannot be considered usable at the Internet layer ("Link Up") until completion of the authenticated key exchange.

However, while [IEEE-802.11i] requires sending of authenticated frames in order to obtain a "Link Up" indication, it does not support management frame authentication. This weakness can be exploited by attackers to enable denial-of-service attacks on stations attached to distant Access Points (APs).

In [IEEE-802.11F], "Link Up" is considered to occur when an AP sends a Reassociation Response. At that point, the AP sends a spoofed frame with the station's source address to a multicast address, thereby causing switches within the Distribution System (DS) to learn the station's MAC address. While this enables forwarding of frames to the station at the new point of attachment, it also permits an attacker to disassociate a station located anywhere within the ESS, by sending an unauthenticated Reassociation Request frame.

4.2. Indication Validation

"Fault Isolation and Recovery" [RFC816], Section 3, describes how hosts interact with routers for the purpose of fault recovery:

Since the gateways always attempt to have a consistent and correct model of the internetwork topology, the host strategy for fault recovery is very simple. Whenever the host feels that something is wrong, it asks the gateway for advice, and, assuming the advice is forthcoming, it believes the advice completely. The advice will be wrong only during the transient period of negotiation, which immediately follows an outage, but will otherwise be reliably correct.

In fact, it is never necessary for a host to explicitly ask a gateway for advice, because the gateway will provide it as appropriate. When a host sends a datagram to some distant net, the host should be prepared to receive back either of two advisory messages which the gateway may send. The ICMP "redirect" message indicates that the gateway to which the host sent the datagram is no longer the best gateway to reach the net in question. The gateway will have forwarded the datagram, but the host should revise its routing table to have a different immediate address for this net. The ICMP "destination unreachable" message indicates that as a result of an outage, it is currently impossible to reach the addressed net or host

in any manner. On receipt of this message, a host can either abandon the connection immediately without any further retransmission, or resend slowly to see if the fault is corrected in reasonable time.

Given today's security environment, it is inadvisable for hosts to act on indications provided by routers without careful consideration. As noted in "ICMP attacks against TCP" [Gont], existing ICMP error messages may be exploited by attackers in order to abort connections in progress, prevent setup of new connections, or reduce throughput of ongoing connections. Similar attacks may also be launched against the Internet layer via forging of ICMP redirects.

Proposals for transported link indications need to demonstrate that they will not add a new set of similar vulnerabilities. Since transported link indications are typically unauthenticated, hosts receiving them may not be able to determine whether they are authentic, or even plausible.

Where link indication proposals may respond to unauthenticated link layer frames, they should utilize upper-layer security mechanisms, where possible. For example, even though a host might utilize an unauthenticated link layer control frame to conclude that a link has become operational, it can use SEND [RFC3971] or authenticated DHCP [RFC3118] in order to obtain secure Internet layer configuration.

4.3. Denial of Service

Link indication proposals need to be particularly careful to avoid enabling denial-of-service attacks that can be mounted at a distance. While wireless links are naturally vulnerable to interference, such attacks can only be perpetrated by an attacker capable of establishing radio contact with the target network. However, attacks that can be mounted from a distance, either by an attacker on another point of attachment within the same network or by an off-link attacker, expand the level of vulnerability.

The transport of link indications can increase risk by enabling vulnerabilities exploitable only by attackers on the local link to be executed across the Internet. Similarly, by integrating link indications with upper layers, proposals may enable a spoofed link layer frame to consume more resources on the host than might otherwise be the case. As a result, while it is important for upper layers to validate link indications, they should not expend excessive resources in doing so.

Congestion control is not only a transport issue, it is also a security issue. In order to not provide leverage to an attacker, a single forged link layer frame should not elicit a magnified response

from one or more hosts, by generating either multiple responses or a single larger response. For example, proposals should not enable multiple hosts to respond to a frame with a multicast destination address.

5. References

5.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

5.2. Informative References

[RFC816] Clark, D., "Fault Isolation and Recovery", RFC 816, July 1982.

[RFC1058] Hedrick, C., "Routing Information Protocol", RFC 1058, June 1988.

[RFC1122] Braden, R., "Requirements for Internet Hosts -- Communication Layers", STD 3, RFC 1122, October 1989.

[RFC1131] Moy, J., "The OSPF Specification", RFC 1131, October 1989.

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, November 1990.

[RFC1256] Deering, S., "ICMP Router Discovery Messages", RFC 1256, September 1991.

[RFC1305] Mills, D., "Network Time Protocol (Version 3) Specification, Implementation and Analysis", RFC 1305, March 1992.

[RFC1307] Young, J. and A. Nicholson, "Dynamically Switched Link Control Protocol", RFC 1307, March 1992.

[RFC1661] Simpson, W., "The Point-to-Point Protocol (PPP)", STD 51, RFC 1661, July 1994.

[RFC1812] Baker, F., "Requirements for IP Version 4 Routers", RFC 1812, June 1995.

[RFC1918] Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, D., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.

[RFC1981] McCann, J., Deering, S. and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, June 1996.

[RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, March 1997.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[RFC2461] Narten, T., Nordmark, E., and W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)", RFC 2461, December 1998.

[RFC2778] Day, M., Rosenberg, J., and H. Sugano, "A Model for Presence and Instant Messaging", RFC 2778, February 2000.

[RFC2861] Handley, M., Padhye, J., and S. Floyd, "TCP Congestion Window Validation", RFC 2861, June 2000.

[RFC2914] Floyd, S., "Congestion Control Principles", RFC 2914, BCP 41, September 2000.

[RFC2923] Lahey, K., "TCP Problems with Path MTU Discovery", RFC 2923, September 2000.

[RFC2960] Stewart, R., Xie, Q., Morneault, K., Sharp, C., Schwarzbauer, H. Taylor, T., Rytina, I., Kalla, M., Zhang, L., and V. Paxson, "Stream Control Transmission Protocol" RFC 2960, October 2000.

[RFC3118] Droms, R. and B. Arbaugh, "Authentication for DHCP Messages", RFC 3118, June 2001.

[RFC3315] Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, July 2003.

[RFC3366] Fairhurst, G. and L. Wood, "Advice to link designers on link Automatic Repeat reQuest (ARQ)", BCP 62, RFC 3366, August 2002.

[RFC3428] Campbell, B., Rosenberg, J., Schulzrinne, H., Huitema, C., and D. Gurle, "Session Initiation Protocol (SIP) Extension for Instant Messaging", RFC 3428, December 2002.

[RFC3748] Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., and H. Levkowetz, "Extensible Authentication Protocol (EAP)", RFC 3748, June 2004.

[RFC3775] Johnson, D., Perkins, C., and J. Arkko, "Mobility Support in IPv6", RFC 3775, June 2004.

[RFC3921] Saint-Andre, P., "Extensible Messaging and Presence protocol (XMPP): Instant Messaging and Presence", RFC 3921, October 2004.

[RFC3927] Cheshire, S., Aboba, B., and E. Guttman, "Dynamic Configuration of Link-Local IPv4 Addresses", RFC 3927, May 2005.

[RFC3971] Arkko, J., Kempf, J., Zill, B., and P. Nikander, "SEcure Neighbor Discovery (SEND)", RFC 3971, March 2005.

[RFC4340] Kohler, E., Handley, M., and S. Floyd, "Datagram Congestion Control Protocol (DCCP)", RFC 4340, March 2006.

[RFC4423] Moskowitz, R. and P. Nikander, "Host Identity Protocol (HIP) Architecture", RFC 4423, May 2006.

[RFC4429] Moore, N., "Optimistic Duplicate Address Detection (DAD) for IPv6", RFC 4429, April 2006.

[RFC4436] Aboba, B., Carlson, J., and S. Cheshire, "Detecting Network Attachment in IPv4 (DNAv4)", RFC 4436, March 2006.

[RFC4821] Mathis, M. and J. Heffner, "Packetization Layer Path MTU Discovery", RFC 4821, March 2007.

[Alimian] Alimian, A., "Roaming Interval Measurements", 11-04-0378-00-roaming-intervals-measurements.ppt, IEEE 802.11 submission (work in progress), March 2004.

[Aguayo] Aguayo, D., Bicket, J., Biswas, S., Judd, G., and R. Morris, "Link-level Measurements from an 802.11b Mesh Network", SIGCOMM '04, September 2004, Portland, Oregon.

[Bakshi] Bakshi, B., Krishna, P., Vadiya, N., and D.Pradhan, "Improving Performance of TCP over Wireless Networks", Proceedings of the 1997 International Conference on Distributed Computer Systems, Baltimore, May 1997.

[BFD] Katz, D. and D. Ward, "Bidirectional Forwarding Detection", Work in Progress, March 2007.

[Biaz] Biaz, S. and N. Vaidya, "Discriminating Congestion Losses from Wireless Losses Using Interarrival Times at the Receiver", Proceedings of the IEEE Symposium on Application-Specific Systems and Software Engineering and Technology, Richardson, TX, Mar 1999.

[CARA] Kim, J., Kim, S., and S. Choi, "CARA: Collision-Aware Rate Adaptation for IEEE 802.11 WLANs", Korean Institute of Communication Sciences (KICS) Journal, Feb. 2006

[Chandran] Chandran, K., Raghunathan, S., Venkatesan, S., and R. Prakash, "A Feedback-Based Scheme for Improving TCP Performance in Ad-Hoc Wireless Networks", Proceedings of the 18th International Conference on Distributed Computing Systems (ICDCS), Amsterdam, May 1998.

[DNAv6] Narayanan, S., "Detecting Network Attachment in IPv6 (DNAv6)", Work in Progress, March 2007.

[E2ELinkup] Dawkins, S. and C. Williams, "End-to-end, Implicit 'Link-Up' Notification", Work in Progress, October 2003.

[EAPIKEv2] Tschofenig, H., Kroeselberg, D., Pashalidis, A., Ohba, Y., and F. Bersani, "EAP IKEv2 Method", Work in Progress, March 2007.

[Eckhardt] Eckhardt, D. and P. Steenkiste, "Measurement and Analysis of the Error Characteristics of an In-Building Wireless Network", SIGCOMM '96, August 1996, Stanford, CA.

[Eddy] Eddy, W. and Y. Swami, "Adapting End Host Congestion Control for Mobility", Technical Report CR-2005-213838, NASA Glenn Research Center, July 2005.

[EfficientEthernet]
Gunaratne, C. and K. Christensen, "Ethernet Adaptive Link Rate: System Design and Performance Evaluation", Proceedings of the IEEE Conference on Local Computer Networks, pp. 28-35, November 2006.

[Eggert] Eggert, L., Schuetz, S., and S. Schmid, "TCP Extensions for Immediate Retransmissions", Work in Progress, June 2005.

[Eggert2] Eggert, L. and W. Eddy, "Towards More Expressive Transport-Layer Interfaces", MobiArch '06, San Francisco, CA.

[ETX] Douglas S. J. De Couto, Daniel Aguayo, John Bicket, and Robert Morris, "A High-Throughput Path Metric for Multi-Hop Wireless Routing", Proceedings of the 9th ACM International Conference on Mobile Computing and Networking (MobiCom '03), San Diego, California, September 2003.

[ETX-Rate] Padhye, J., Draves, R. and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks", Proceedings of ACM MobiCom Conference, September 2003.

[ETX-Radio] Kulkarni, G., Nandan, A., Gerla, M., and M. Srivastava, "A Radio Aware Routing Protocol for Wireless Mesh Networks", UCLA Computer Science Department, Los Angeles, CA.

[GenTrig] Gupta, V. and D. Johnston, "A Generalized Model for Link Layer Triggers", submission to IEEE 802.21 (work in progress), March 2004, available at: <http://www.ieee802.org/handoff/march04_meeting_docs/Generalized_triggers-02.pdf>.

[Goel] Goel, S. and D. Sanghi, "Improving TCP Performance over Wireless Links", Proceedings of TENCON'98, pages 332-335. IEEE, December 1998.

[Gont] Gont, F., "ICMP attacks against TCP", Work in Progress, October 2006.

[Gurtov] Gurtov, A. and J. Korhonen, "Effect of Vertical Handovers on Performance of TCP-Friendly Rate Control", to appear in ACM MCCR, 2004.

[GurtovFloyd] Gurtov, A. and S. Floyd, "Modeling Wireless Links for Transport Protocols", Computer Communications Review (CCR) 34, 2 (2003).

[Haratcherev] Haratcherev, I., Lagendijk, R., Langendoen, K., and H. Sips, "Hybrid Rate Control for IEEE 802.11", MobiWac '04, October 1, 2004, Philadelphia, Pennsylvania, USA.

[Haratcherev2] Haratcherev, I., "Application-oriented Link Adaptation for IEEE 802.11", Ph.D. Thesis, Technical University of Delft, Netherlands, ISBN-10:90-9020513-6, ISBN-13:978-90-9020513-7, March 2006.

[HMP] Lee, S., Cho, J., and A. Campbell, "Hotspot Mitigation Protocol (HMP)", Work in Progress, October 2003.

[Holland] Holland, G. and N. Vaidya, "Analysis of TCP Performance over Mobile Ad Hoc Networks", Proceedings of the Fifth International Conference on Mobile Computing and Networking, pages 219-230. ACM/IEEE, Seattle, August 1999.

[Iannaccone] Iannaccone, G., Chuah, C., Mortier, R., Bhattacharyya, S., and C. Diot, "Analysis of link failures in an IP backbone", Proc. of ACM Sigcomm Internet Measurement Workshop, November, 2002.

[IEEE-802.1X] Institute of Electrical and Electronics Engineers, "Local and Metropolitan Area Networks: Port-Based Network Access Control", IEEE Standard 802.1X, December 2004.

[IEEE-802.11] Institute of Electrical and Electronics Engineers, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications", IEEE Standard 802.11, 2003.

[IEEE-802.11e] Institute of Electrical and Electronics Engineers, "Standard for Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements", IEEE 802.11e, November 2005.

[IEEE-802.11F] Institute of Electrical and Electronics Engineers, "IEEE Trial-Use Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation", IEEE 802.11F, June 2003 (now deprecated).

[IEEE-802.11i] Institute of Electrical and Electronics Engineers, "Supplement to Standard for Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Specification for Enhanced Security", IEEE 802.11i, July 2004.

[IEEE-802.11k] Institute of Electrical and Electronics Engineers, "Draft Amendment to Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 7: Radio Resource Management", IEEE 802.11k/D7.0, January 2007.

[IEEE-802.21] Institute of Electrical and Electronics Engineers, "Draft Standard for Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 21: Media Independent Handover", IEEE 802.21D0, June 2005.

[Kamerman] Kamerman, A. and L. Monteban, "WaveLAN II: A High-Performance Wireless LAN for the Unlicensed Band", Bell Labs Technical Journal, Summer 1997.

[Kim] Kim, K., Park, Y., Suh, K., and Y. Park, "The BU-trigger method for improving TCP performance over Mobile IPv6", Work in Progress, August 2004.

[Kotz] Kotz, D., Newport, C., and C. Elliot, "The mistaken axioms of wireless-network research", Dartmouth College Computer Science Technical Report TR2003-467, July 2003.

[Krishnan] Krishnan, R., Sterbenz, J., Eddy, W., Partridge, C., and M. Allman, "Explicit Transport Error Notification (ETEN) for Error-Prone Wireless and Satellite Networks", Computer Networks, 46 (3), October 2004.

[Lacage] Lacage, M., Manshaei, M., and T. Turletti, "IEEE 802.11 Rate Adaptation: A Practical Approach", MSWiM '04, October 4-6, 2004, Venezia, Italy.

[Lee] Park, S., Lee, M., and J. Korhonen, "Link Characteristics Information for Mobile IP", Work in Progress, January 2007.

[Ludwig] Ludwig, R. and B. Rathonyi, "Link-layer Enhancements for TCP/IP over GSM", Proceedings of IEEE Infocom '99, March 1999.

[MIPEAP] Giaretta, C., Guardini, I., Demaria, E., Bournelle, J., and M. Laurent-Maknavicius, "MIPv6 Authorization and Configuration based on EAP", Work in Progress, October 2006.

[Mishra] Mitra, A., Shin, M., and W. Arbaugh, "An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process", CS-TR-4395, University of Maryland Department of Computer Science, September 2002.

[Morgan] Morgan, S. and S. Keshav, "Packet-Pair Rate Control - Buffer Requirements and Overload Performance", Technical Memorandum, AT&T Bell Laboratories, October 1994.

[Mun] Mun, Y. and J. Park, "Layer 2 Handoff for Mobile-IPv4 with 802.11", Work in Progress, March 2004.

[ONOE] Onoe Rate Control, <http://madwifi.org/browser/trunk/ath_rate/onoe>.

[Park] Park, S., Njedjou, E., and N. Montavont, "L2 Triggers Optimized Mobile IPv6 Vertical Handover: The 802.11/GPRS Example", Work in Progress, July 2004.

[Pavon] Pavon, J. and S. Choi, "Link adaptation strategy for IEEE802.11 WLAN via received signal strength measurement", IEEE International Conference on Communications, 2003 (ICC '03), volume 2, pages 1108-1113, Anchorage, Alaska, USA, May 2003.

[PEAP] Palekar, A., Simon, D., Salowey, J., Zhou, H., Zorn, G., and S. Josefsson, "Protected EAP Protocol (PEAP) Version 2", Work in Progress, October 2004.

[PRNET] Jubin, J. and J. Tornow, "The DARPA packet radio network protocols", Proceedings of the IEEE, 75(1), January 1987.

[Qiao] Qiao D., Choi, S., Jain, A., and Kang G. Shin, "MiSer: An Optimal Low-Energy Transmission Strategy for IEEE 802.11 a/h", in Proc. ACM MobiCom'03, San Diego, CA, September 2003.

[RBAR] Holland, G., Vaidya, N., and P. Bahl, "A Rate-Adaptive MAC Protocol for Multi-Hop Wireless Networks", Proceedings ACM MOBICOM, July 2001.

[Ramani] Ramani, I. and S. Savage, "SyncScan: Practical Fast Handoff for 802.11 Infrastructure Networks", Proceedings of the IEEE InfoCon 2005, March 2005.

[Robust] Wong, S., Yang, H ., Lu, S., and V. Bharghavan, "Robust Rate Adaptation for 802.11 Wireless Networks", ACM MobiCom'06, Los Angeles, CA, September 2006.

[SampleRate] Bicket, J., "Bit-rate Selection in Wireless networks", MIT Master's Thesis, 2005.

[Scott] Scott, J., Mapp, G., "Link Layer Based TCP Optimisation for Disconnecting Networks", ACM SIGCOMM Computer Communication Review, 33(5), October 2003.

[Schuetz] Schutz, S., Eggert, L., Schmid, S., and M. Brunner, "Protocol Enhancements for Intermittently Connected Hosts", ACM SIGCOMM Computer Communications Review, Volume 35, Number 2, July 2005.

[Shortest] Douglas S. J. De Couto, Daniel Aguayo, Benjamin A. Chambers and Robert Morris, "Performance of Multihop Wireless Networks: Shortest Path is Not Enough", Proceedings of the First Workshop on Hot Topics in Networking (HotNets-I), Princeton, New Jersey, October 2002.

[TRIGTRAN] Dawkins, S., Williams, C., and A. Yegin, "Framework and Requirements for TRIGTRAN", Work in Progress, August 2003.

[Vatn] Vatn, J., "An experimental study of IEEE 802.11b handover performance and its effect on voice traffic", TRITA-IMIT-TSLAB R 03:01, KTH Royal Institute of Technology, Stockholm, Sweden, July 2003.

[Velayos] Velayos, H. and G. Karlsson, "Techniques to Reduce IEEE 802.11b MAC Layer Handover Time", TRITA-IMIT-LCN R 03:02, KTH Royal Institute of Technology, Stockholm, Sweden, April 2003.

[Vertical] Zhang, Q., Guo, C., Guo, Z., and W. Zhu, "Efficient Mobility Management for Vertical Handoff between WWAN and WLAN", IEEE Communications Magazine, November 2003.

[Villamizar] Villamizar, C., "OSPF Optimized Multipath (OSPF-OMP)", Work in Progress, February 1999.

[Xylomenos] Xylomenos, G., "Multi Service Link Layers: An Approach to Enhancing Internet Performance over Wireless Links", Ph.D. thesis, University of California at San Diego, 1999.

[Yegin] Yegin, A., "Link-layer Triggers Protocol", Work in Progress, June 2002.

6. Acknowledgments

The authors would like to acknowledge James Kempf, Phil Roberts, Gorry Fairhurst, John Wroclawski, Aaron Falk, Sally Floyd, Pekka Savola, Pekka Nikander, Dave Thaler, Yogesh Swami, Wesley Eddy, and Janne Peisa for contributions to this document.

Appendix A. Literature Review

This appendix summarizes the literature with respect to link indications on wireless local area networks.

A.1. Link Layer

The characteristics of wireless links have been found to vary considerably depending on the environment.

In "Performance of Multihop Wireless Networks: Shortest Path is Not Enough" [Shortest], the authors studied the performance of both an indoor and outdoor mesh network. By measuring inter-node throughput, the best path between nodes was computed. The throughput of the best path was compared with the throughput of the shortest path computed based on a hop-count metric. In almost all cases, the shortest path route offered considerably lower throughput than the best path.

In examining link behavior, the authors found that rather than exhibiting a bi-modal distribution between "up" (low loss rate) and "down" (high loss rate), many links exhibited intermediate loss rates. Asymmetry was also common, with 30 percent of links demonstrating substantial differences in the loss rates in each direction. As a result, on wireless networks the measured throughput can differ substantially from the negotiated rate due to retransmissions, and successful delivery of routing packets is not necessarily an indication that the link is useful for delivery of data.

In "Measurement and Analysis of the Error Characteristics of an In-Building Wireless Network" [Eckhardt], the authors characterize the performance of an AT&T Wavelan 2 Mbps in-building WLAN operating in Infrastructure mode on the Carnegie Mellon campus. In this study, very low frame loss was experienced. As a result, links could be assumed to operate either very well or not at all.

In "Link-level Measurements from an 802.11b Mesh Network" [Aguayo], the authors analyze the causes of frame loss in a 38-node urban multi-hop 802.11 ad-hoc network. In most cases, links that are very bad in one direction tend to be bad in both directions, and links that are very good in one direction tend to be good in both directions. However, 30 percent of links exhibited loss rates differing substantially in each direction.

Signal to noise ratio (SNR) and distance showed little value in predicting loss rates, and rather than exhibiting a step-function transition between "up" (low loss) or "down" (high loss) states, inter-node loss rates varied widely, demonstrating a nearly uniform

distribution over the range at the lower rates. The authors attribute the observed effects to multi-path fading, rather than attenuation or interference.

The findings of [Eckhardt] and [Aguayo] demonstrate the diversity of link conditions observed in practice. While for indoor infrastructure networks site surveys and careful measurement can assist in promoting ideal behavior, in ad-hoc/mesh networks node mobility and external factors such as weather may not be easily controlled.

Considerable diversity in behavior is also observed due to implementation effects. "Techniques to reduce IEEE 802.11b MAC layer handover time" [Velayos] measured handover times for a stationary STA after the AP was turned off. This study divided handover times into detection (determination of disconnection from the existing point of attachment), search (discovery of alternative attachment points), and execution (connection to an alternative point of attachment) phases. These measurements indicated that the duration of the detection phase (the largest component of handoff delay) is determined by the number of non-acknowledged frames triggering the search phase and delays due to precursors such as RTS/CTS and rate adaptation.

Detection behavior varied widely between implementations. For example, network interface cards (NICs) designed for desktops attempted more retransmissions prior to triggering search as compared with laptop designs, since they assumed that the AP was always in range, regardless of whether the Beacon was received.

The study recommends that the duration of the detection phase be reduced by initiating the search phase as soon as collisions can be excluded as the cause of non-acknowledged transmissions; the authors recommend three consecutive transmission failures as the cutoff. This approach is both quicker and more immune to multi-path interference than monitoring of the SNR. Where the STA is not sending or receiving frames, it is recommended that Beacon reception be tracked in order to detect disconnection, and that Beacon spacing be reduced to 60 ms in order to reduce detection times. In order to compensate for more frequent triggering of the search phase, the authors recommend algorithms for wait time reduction, as well as interleaving of search and data frame transmission.

"An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process" [Mishra] investigates handoff latencies obtained with three mobile STA implementations communicating with two APs. The study found that there is a large variation in handoff latency among STA and AP implementations and that implementations utilize different message sequences. For example, one STA sends a Reassociation Request prior

to authentication, which results in receipt of a Deauthenticate message. The study divided handoff latency into discovery, authentication, and reassociation exchanges, concluding that the discovery phase was the dominant component of handoff delay. Latency in the detection phase was not investigated.

"SyncScan: Practical Fast Handoff for 802.11 Infrastructure Networks" [Ramani] weighs the pros and cons of active versus passive scanning. The authors point out the advantages of timed Beacon reception, which had previously been incorporated into [IEEE-802.11k]. Timed Beacon reception allows the station to continually keep up to date on the signal to noise ratio of neighboring APs, allowing handoff to occur earlier. Since the station does not need to wait for initial and subsequent responses to a broadcast Probe Response (MinChannelTime and MaxChannelTime, respectively), performance is comparable to what is achievable with 802.11k Neighbor Reports and unicast Probe Requests.

The authors measured the channel switching delay, the time it takes to switch to a new frequency and begin receiving frames. Measurements ranged from 5 ms to 19 ms per channel; where timed Beacon reception or interleaved active scanning is used, switching time contributes significantly to overall handoff latency. The authors propose deployment of APs with Beacons synchronized via Network Time Protocol (NTP) [RFC1305], enabling a driver implementing SyncScan to work with legacy APs without requiring implementation of new protocols. The authors measured the distribution of inter-arrival times for stations implementing SyncScan, with excellent results.

"Roaming Interval Measurements" [Alimian] presents data on the behavior of stationary STAs after the AP signal has been shut off. This study highlighted implementation differences in rate adaptation as well as detection, scanning, and handoff. As in [Velayos], performance varied widely between implementations, from half an order of magnitude variation in rate adaptation to an order of magnitude difference in detection times, two orders of magnitude in scanning, and one and a half orders of magnitude in handoff times.

"An experimental study of IEEE 802.11b handoff performance and its effect on voice traffic" [Vatn] describes handover behavior observed when the signal from the AP is gradually attenuated, which is more representative of field experience than the shutoff techniques used in [Velayos]. Stations were configured to initiate handover when signal strength dipped below a threshold, rather than purely based on frame loss, so that they could begin handover while still connected to the current AP. It was noted that stations continued to receive data frames during the search phase. Station-initiated

Disassociation and pre-authentication were not observed in this study.

A.1.1. Link Indications

Within a link layer, the definition of "Link Up" and "Link Down" may vary according to the deployment scenario. For example, within PPP [RFC1661], either peer may send an LCP-Terminate frame in order to terminate the PPP link layer, and a link may only be assumed to be usable for sending network protocol packets once Network Control Protocol (NCP) negotiation has completed for that protocol.

Unlike PPP, IEEE 802 does not include facilities for network layer configuration, and the definition of "Link Up" and "Link Down" varies by implementation. Empirical evidence suggests that the definition of "Link Up" and "Link Down" may depend on whether the station is mobile or stationary, whether infrastructure or ad-hoc mode is in use, and whether security and Inter-Access Point Protocol (IAPP) is implemented.

Where a STA encounters a series of consecutive non-acknowledged frames while having missed one or more Beacons, the most likely cause is that the station has moved out of range of the AP. As a result, [Velayos] recommends that the station begin the search phase after collisions can be ruled out; since this approach does not take rate adaptation into account, it may be somewhat aggressive. Only when no alternative workable rate or point of attachment is found is a "Link Down" indication returned.

In a stationary point-to-point installation, the most likely cause of an outage is that the link has become impaired, and alternative points of attachment may not be available. As a result, implementations configured to operate in this mode tend to be more persistent. For example, within 802.11 the short interframe space (SIFS) interval may be increased and MIB variables relating to timeouts (such as dot11AuthenticationResponseTimeout, dot11AssociationResponseTimeout, dot11ShortRetryLimit, and dot11LongRetryLimit) may be set to larger values. In addition, a "Link Down" indication may be returned later.

In IEEE 802.11 ad-hoc mode with no security, reception of data frames is enabled in State 1 ("Unauthenticated" and "Unassociated"). As a result, reception of data frames is enabled at any time, and no explicit "Link Up" indication exists.

In Infrastructure mode, IEEE 802.11-2003 enables reception of data frames only in State 3 ("Authenticated" and "Associated"). As a result, a transition to State 3 (e.g., completion of a successful

Association or Reassociation exchange) enables sending and receiving of network protocol packets and a transition from State 3 to State 2 (reception of a "Disassociate" frame) or State 1 (reception of a "Deauthenticate" frame) disables sending and receiving of network protocol packets. As a result, IEEE 802.11 stations typically signal "Link Up" on receipt of a successful Association/Reassociation Response.

As described within [IEEE-802.11F], after sending a Reassociation Response, an Access Point will send a frame with the station's source address to a multicast destination. This causes switches within the Distribution System (DS) to update their learning tables, readying the DS to forward frames to the station at its new point of attachment. Were the AP to not send this "spoofed" frame, the station's location would not be updated within the distribution system until it sends its first frame at the new location. Thus, the purpose of spoofing is to equalize uplink and downlink handover times. This enables an attacker to deny service to authenticated and associated stations by spoofing a Reassociation Request using the victim's MAC address, from anywhere within the ESS. Without spoofing, such an attack would only be able to disassociate stations on the AP to which the Reassociation Request was sent.

The signaling of "Link Down" is considerably more complex. Even though a transition to State 2 or State 1 results in the station being unable to send or receive IP packets, this does not necessarily imply that such a transition should be considered a "Link Down" indication. In an infrastructure network, a station may have a choice of multiple Access Points offering connection to the same network. In such an environment, a station that is unable to reach State 3 with one Access Point may instead choose to attach to another Access Point. Rather than registering a "Link Down" indication with each move, the station may instead register a series of "Link Up" indications.

In [IEEE-802.11i], forwarding of frames from the station to the distribution system is only feasible after the completion of the 4-way handshake and group-key handshake, so that entering State 3 is no longer sufficient. This has resulted in several observed problems. For example, where a "Link Up" indication is triggered on the station by receipt of an Association/Reassociation Response, DHCP [RFC2131] or Router Solicitation/Router Advertisement (RS/RA) may be triggered prior to when the link is usable by the Internet layer, resulting in configuration delays or failures. Similarly, transport layer connections will encounter packet loss, resulting in back-off of retransmission timers.

A.1.2. Smart Link Layer Proposals

In order to improve link layer performance, several studies have investigated "smart link layer" proposals.

"Advice to link designers on link Automatic Repeat reQuest (ARQ)" [RFC3366] provides advice to the designers of digital communication equipment and link-layer protocols employing link-layer Automatic Repeat reQuest (ARQ) techniques for IP. It discusses the use of ARQ, timers, persistency in retransmission, and the challenges that arise from sharing links between multiple flows and from different transport requirements.

In "Link-layer Enhancements for TCP/IP over GSM" [Ludwig], the authors describe how the Global System for Mobile Communications (GSM)-reliable and unreliable link layer modes can be simultaneously utilized without higher layer control. Where a reliable link layer protocol is required (where reliable transports such TCP and Stream Control Transmission Protocol (SCTP) [RFC2960] are used), the Radio Link Protocol (RLP) can be engaged; with delay-sensitive applications such as those based on UDP, the transparent mode (no RLP) can be used. The authors also describe how PPP negotiation can be optimized over high-latency GSM links using "Quickstart-PPP".

In "Link Layer Based TCP Optimisation for Disconnecting Networks" [Scott], the authors describe performance problems that occur with reliable transport protocols facing periodic network disconnections, such as those due to signal fading or handoff. The authors define a disconnection as a period of connectivity loss that exceeds a retransmission timeout, but is shorter than the connection lifetime. One issue is that link-unaware senders continue to back off during periods of disconnection. The authors suggest that a link-aware reliable transport implementation halt retransmission after receiving a "Link Down" indication. Another issue is that on reconnection the lengthened retransmission times cause delays in utilizing the link.

To improve performance, a "smart link layer" is proposed, which stores the first packet that was not successfully transmitted on a connection, then retransmits it upon receipt of a "Link Up" indication. Since a disconnection can result in hosts experiencing different network conditions upon reconnection, the authors do not advocate bypassing slow start or attempting to raise the congestion window. Where IPsec is used and connections cannot be differentiated because transport headers are not visible, the first untransmitted packet for a given sender and destination IP address can be retransmitted. In addition to looking at retransmission of a single packet per connection, the authors also examined other schemes such

as retransmission of multiple packets and simulated duplicate reception of single or multiple packets (known as rereception).

In general, retransmission schemes were superior to rereception schemes, since rereception cannot stimulate fast retransmit after a timeout. Retransmission of multiple packets did not appreciably improve performance over retransmission of a single packet. Since the focus of the research was on disconnection rather than just lossy channels, a two-state Markov model was used, with the "up" state representing no loss, and the "down" state representing 100 percent loss.

In "Multi Service Link Layers: An Approach to Enhancing Internet Performance over Wireless Links" [Xylomenos], the authors use ns-2 to simulate the performance of various link layer recovery schemes (raw link without retransmission, go back N, XOR-based FEC, selective repeat, Karn's RLP, out-of-sequence RLP, and Berkeley Snoop) in stand-alone file transfer, Web browsing, and continuous media distribution. While selective repeat and Karn's RLP provide the highest throughput for file transfer and Web browsing scenarios, continuous media distribution requires a combination of low delay and low loss and the out-of-sequence RLP performed best in this scenario. Since the results indicate that no single link layer recovery scheme is optimal for all applications, the authors propose that the link layer implement multiple recovery schemes. Simulations of the multi-service architecture showed that the combination of a low-error rate recovery scheme for TCP (such as Karn's RLP) and a low-delay scheme for UDP traffic (such as out-of-sequence RLP) provides for good performance in all scenarios. The authors then describe how a multi-service link layer can be integrated with Differentiated Services.

In "WaveLAN-II: A High-Performance Wireless LAN for the Unlicensed Band" [Kamerman], the authors propose an open-loop rate adaptation algorithm known as Automatic Rate Fallback (ARF). In ARF, the sender adjusts the rate upwards after a fixed number of successful transmissions, and adjusts the rate downwards after one or two consecutive failures. If after an upwards rate adjustment the transmission fails, the rate is immediately readjusted downwards.

In "A Rate-Adaptive MAC Protocol for Multi-Hop Wireless Networks" [RBAR], the authors propose a closed-loop rate adaptation approach that requires incompatible changes to the IEEE 802.11 MAC. In order to enable the sender to better determine the transmission rate, the receiver determines the packet length and signal to noise ratio (SNR) of a received RTS frame and calculates the corresponding rate based on a theoretical channel model, rather than channel usage statistics. The recommended rate is sent back in the CTS frame. This allows the

rate (and potentially the transmit power) to be optimized on each transmission, albeit at the cost of requiring RTS/CTS for every frame transmission.

In "MiSer: An Optimal Low-Energy Transmission Strategy for IEEE 802.11 a/h" [Qiao], the authors propose a scheme for optimizing transmit power. The proposal mandates the use of RTS/CTS in order to deal with hidden nodes, requiring that CTS and ACK frames be sent at full power. The authors utilize a theoretical channel model rather than one based on channel usage statistics.

In "IEEE 802.11 Rate Adaptation: A Practical Approach" [Lacage], the authors distinguish between low-latency implementations, which enable per-packet rate decisions, and high-latency implementations, which do not. The former implementations typically include dedicated CPUs in their design, enabling them to meet real-time requirements. The latter implementations are typically based on highly integrated designs in which the upper MAC is implemented on the host. As a result, due to operating system latencies the information required to make per-packet rate decisions may not be available in time.

The authors propose an Adaptive ARF (AARF) algorithm for use with low-latency implementations. This enables rapid downward rate negotiation on failure to receive an ACK, while increasing the number of successful transmissions required for upward rate negotiation. The AARF algorithm is therefore highly stable in situations where channel properties are changing slowly, but slow to adapt upwards when channel conditions improve. In order to test the algorithm, the authors utilized ns-2 simulations as well as implementing a version of AARF adapted to a high-latency implementation, the AR 5212 chipset. The Multiband Atheros Driver for WiFi (MadWiFi) driver enables a fixed schedule of rates and retries to be provided when a frame is queued for transmission. The adapted algorithm, known as the Adaptive Multi Rate Retry (AMRR), requests only one transmission at each of three rates, the last of which is the minimum available rate. This enables adaptation to short-term fluctuations in the channel with minimal latency. The AMRR algorithm provides performance considerably better than the existing MadWifi driver.

In "Link Adaptation Strategy for IEEE 802.11 WLAN via Received Signal Strength Measurement" [Pavon], the authors propose an algorithm by which a STA adjusts the transmission rate based on a comparison of the received signal strength (RSS) from the AP with dynamically estimated threshold values for each transmission rate. Upon reception of a frame, the STA updates the average RSS, and on transmission the STA selects a rate and adjusts the RSS threshold values based on whether or not the transmission is successful. In order to validate the algorithm, the authors utilized an OPNET

simulation without interference, and an ideal curve of bit error rate (BER) vs. signal to noise ratio (SNR) was assumed. Not surprisingly, the simulation results closely matched the maximum throughput achievable for a given signal to noise ratio, based on the ideal BER vs. SNR curve.

In "Hybrid Rate Control for IEEE 802.11" [Haratcherev], the authors describe a hybrid technique utilizing Signal Strength Indication (SSI) data to constrain the potential rates selected by statistics-based automatic rate control. Statistics-based rate control techniques include:

Maximum Throughput

This technique, which was chosen as the statistics-based technique in the hybrid scheme, sends a fraction of data at adjacent rates in order to estimate which rate provides the maximum throughput. Since accurate estimation of throughput requires a minimum number of frames to be sent at each rate, and only a fraction of frames are utilized for this purpose, this technique adapts more slowly at lower rates; with 802.11b rates, the adaptation time scale is typically on the order of a second. Depending on how many rates are tested, this technique can enable adaptation beyond adjacent rates. However, where maximum rate and low frame loss are already being encountered, this technique results in lower throughput.

Frame Error Rate (FER) Control

This technique estimates the FER, attempting to keep it between a lower limit (if FER moves below, increase rate) and upper limit (if FER moves above, decrease rate). Since this technique can utilize all the transmitted data, it can respond faster than maximum throughput techniques. However, there is a tradeoff of reaction time versus FER estimation accuracy; at lower rates either reaction times slow or FER estimation accuracy will suffer. Since this technique only measures the FER at the current rate, it can only enable adaptation to adjacent rates.

Retry-based

This technique modifies FER control techniques by enabling rapid downward rate adaptation after a number (5-10) of unsuccessful retransmissions. Since fewer packets are required, the sensitivity of reaction time to rate is reduced. However, upward rate adaptation proceeds more slowly since it is based on a collection of FER data. This technique is limited to adaptation to adjacent rates, and it has the disadvantage of potentially worsening frame loss due to contention.

While statistics-based techniques are robust against short-lived link quality changes, they do not respond quickly to long-lived changes. By constraining the rate selected by statistics-based techniques based on ACK SSI versus rate data (not theoretical curves), more rapid link adaptation was enabled. In order to ensure rapid adaptation during rapidly varying conditions, the rate constraints are tightened when the SSI values are changing rapidly, encouraging rate transitions. The authors validated their algorithms by implementing a driver for the Atheros AR5000 chipset, and then testing its response to insertion and removal from a microwave oven acting as a Faraday cage. The hybrid algorithm dropped many fewer packets than the maximum throughput technique by itself.

In order to estimate the SSI of data at the receiver, the ACK SSI was used. This approach does not require the receiver to provide the sender with the received power, so that it can be implemented without changing the IEEE 802.11 MAC. Calibration of the rate versus ACK SSI curves does not require a symmetric channel, but it does require that channel properties in both directions vary in a proportional way and that the ACK transmit power remains constant. The authors checked the proportionality assumption and found that the SSI of received data correlated highly (74%) with the SSI of received ACKs. Low pass filtering and monotonicity constraints were applied to remove noise in the rate versus SSI curves. The resulting hybrid rate adaptation algorithm demonstrated the ability to respond to rapid deterioration (and improvement) in channel properties, since it is not restricted to moving to adjacent rates.

In "CARA: Collision-Aware Rate Adaptation for IEEE 802.11 WLANs" [CARA], the authors propose Collision-Aware Rate Adaptation (CARA). This involves utilization of Clear Channel Assessment (CCA) along with adaptation of the Request-to-Send/Clear-to-Send (RTS/CTS) mechanism to differentiate losses caused by frame collisions from losses caused by channel conditions. Rather than decreasing rate as the result of frame loss due to collisions, which leads to increased contention, CARA selectively enables RTS/CTS (e.g., after a frame loss), reducing the likelihood of frame loss due to hidden stations. CARA can also utilize CCA to determine whether a collision has occurred after a transmission; however, since CCA may not detect a significant fraction of all collisions (particularly when transmitting at low rate), its use is optional. As compared with ARF, in simulations the authors show large improvements in aggregate throughput due to addition of adaptive RTS/CTS, and additional modest improvements with the additional help of CCA.

In "Robust Rate Adaptation for 802.11 Wireless Networks" [Robust], the authors implemented the ARF, AARF, and SampleRate [SampleRate] algorithms on a programmable Access Point platform, and

experimentally examined the performance of these algorithms as well as the ONOE [ONOE] algorithm implemented in MadWiFi. Based on their experiments, the authors critically examine the assumptions underlying existing rate negotiation algorithms:

Decrease transmission rate upon severe frame loss
: Where severe frame loss is due to channel conditions, rate reduction can improve throughput. However, where frame loss is due to contention (such as from hidden stations), reducing transmission rate increases congestion, lowering throughput and potentially leading to congestive collapse. Instead, the authors propose adaptive enabling of RTS/CTS so as to reduce contention due to hidden stations. Once RTS/CTS is enabled, remaining losses are more likely to be due to channel conditions, providing more reliable guidance on increasing or decreasing transmission rate.

Use probe frames to assess possible new rates
: Probe frames reliably estimate frame loss at a given rate unless the sample size is sufficient and the probe frames are of comparable length to data frames. The authors argue that rate adaptation schemes such as SampleRate are too sensitive to loss of probe packets. In order to satisfy sample size constraints, a significant number of probe frames are required. This can increase frame loss if the probed rate is too high, and can lower throughput if the probed rate is too low. Instead, the authors propose assessment of the channel condition by tracking the frame loss ratio within a window of 5 to 40 frames.

Use consecutive transmission successes/losses to increase/decrease rate
: The authors argue that consecutive successes or losses are not a reliable basis for rate increases or decreases; greater sample size is needed.

Use PHY metrics like SNR to infer new transmission rate
: The authors argue that received signal to noise ratio (SNR) routinely varies 5 dB per packet and that variations of 10-14 dB are common. As a result, rate decisions based on SNR or signal strength can cause transmission rate to vary rapidly. The authors question the value of such rapid variation, since studies such as [Aguayo] show little correlation between SNR and frame loss probability. As a result, the authors argue that neither received signal strength indication (RSSI) nor background energy level can be used to distinguish losses due to contention from those due to channel conditions. While multi-path interference can simultaneously result in high signal strength and frame loss, the relationship between low signal

strength and high frame loss is stronger. Therefore, transmission rate decreases due to low received signal strength probably do reflect sudden worsening in channel conditions, although sudden increases may not necessarily indicate that channel conditions have improved.

Long-term smoothened operation produces best average performance
The authors present evidence that frame losses more than 150 ms apart are uncorrelated. Therefore, collection of statistical data over intervals of 1 second or greater reduces responsiveness, but does not improve the quality of transmission rate decisions. Rather, the authors argue that a sampling period of 100 ms provides the best average performance. Such small sampling periods also argue against use of probes, since probe packets can only represent a fraction of all data frames and probes collected more than 150 ms apart may not provide reliable information on channel conditions.

Based on these flaws, the authors propose the Robust Rate Adaptation Algorithm (RRAA). RRAA utilizes only the frame loss ratio at the current transmission rate to determine whether to increase or decrease the transmission rate; PHY layer information or probe packets are not used. Each transmission rate is associated with an estimation window, a maximum tolerable loss threshold (MTL) and an opportunistic rate increase threshold (ORI). If the loss ratio is larger than the MTL, the transmission rate is decreased, and if it is smaller than the ORI, transmission rate is increased; otherwise transmission rate remains the same. The thresholds are selected in order to maximize throughput. Although RRAA only allows movement between adjacent transmission rates, the algorithm does not require collection of an entire estimation window prior to increasing or decreasing transmission rates; if additional data collection would not change the decision, the change is made immediately.

The authors validate the RRAA algorithm using experiments and field trials; the results indicate that RRAA without adaptive RTS/CTS outperforms the ARF, AARF, and Sample Rate algorithms. This occurs because RRAA is not as sensitive to transient frame loss and does not use probing, enabling it to more frequently utilize higher transmission rates. Where there are no hidden stations, turning on adaptive RTS/CTS reduces performance by at most a few percent. However, where there is substantial contention from hidden stations, adaptive RTS/CTS provides large performance gains, due to reduction in frame loss that enables selection of a higher transmission rate.

In "Efficient Mobility Management for Vertical Handoff between WWAN and WLAN" [Vertical], the authors propose use of signal strength and link utilization in order to optimize vertical handoff. WLAN to WWAN

handoff is driven by SSI decay. When IEEE 802.11 SSI falls below a threshold (S1), Fast Fourier Transform (FFT)-based decay detection is undertaken to determine if the signal is likely to continue to decay. If so, then handoff to the WWAN is initiated when the signal falls below the minimum acceptable level (S2). WWAN to WLAN handoff is driven by both PHY and MAC characteristics of the IEEE 802.11 target network. At the PHY layer, characteristics such as SSI are examined to determine if the signal strength is greater than a minimum value (S3). At the MAC layer, the IEEE 802.11 Network Allocation Vector (NAV) occupation is examined in order to estimate the maximum available bandwidth and mean access delay. Note that depending on the value of S3, it is possible for the negotiated rate to be less than the available bandwidth. In order to prevent premature handoff between WLAN and WWAN, S1 and S2 are separated by 6 dB; in order to prevent oscillation between WLAN and WWAN media, S3 needs to be greater than S1 by an appropriate margin.

A.2. Internet Layer

Within the Internet layer, proposals have been made for utilizing link indications to optimize IP configuration, to improve the usefulness of routing metrics, and to optimize aspects of Mobile IP handoff.

In "Analysis of link failures in an IP backbone" [Iannaccone], the authors investigate link failures in Sprint’s IP backbone. They identify the causes of convergence delay, including delays in detection of whether an interface is down or up. While it is fastest for a router to utilize link indications if available, there are situations in which it is necessary to depend on loss of routing packets to determine the state of the link. Once the link state has been determined, a delay may occur within the routing protocol in order to dampen link flaps. Finally, another delay may be introduced in propagating the link state change, in order to rate limit link state advertisements, and guard against instability.

"Bidirectional Forwarding Detection" [BFD] notes that link layers may provide only limited failure indications, and that relatively slow "Hello" mechanisms are used in routing protocols to detect failures when no link layer indications are available. This results in failure detection times of the order of a second, which is too long for some applications. The authors describe a mechanism that can be used for liveness detection over any media, enabling rapid detection of failures in the path between adjacent forwarding engines. A path is declared operational when bidirectional reachability has been confirmed.

In "Detecting Network Attachment (DNA) in IPv4" [RFC4436], a host that has moved to a new point of attachment utilizes a bidirectional reachability test in parallel with DHCP [RFC2131] to rapidly reconfirm an operable configuration.

In "L2 Triggers Optimized Mobile IPv6 Vertical Handover: The 802.11/GPRS Example" [Park], the authors propose that the mobile node send a router solicitation on receipt of a "Link Up" indication in order to provide lower handoff latency than would be possible using generic movement detection [RFC3775]. The authors also suggest immediate invalidation of the Care-of Address (CoA) on receipt of a "Link Down" indication. However, this is problematic where a "Link Down" indication can be followed by a "Link Up" indication without a resulting change in IP configuration, as described in [RFC4436].

In "Layer 2 Handoff for Mobile-IPv4 with 802.11" [Mun], the authors suggest that MIPv4 Registration messages be carried within Information Elements of IEEE 802.11 Association/Reassociation frames, in order to minimize handoff delays. This requires modification to the mobile node as well as 802.11 APs. However, prior to detecting network attachment, it is difficult for the mobile node to determine whether or not the new point of attachment represents a change of network. For example, even where a station remains within the same ESS, it is possible that the network will change. Where no change of network results, sending a MIPv4 Registration message with each Association/Reassociation is unnecessary. Where a change of network results, it is typically not possible for the mobile node to anticipate its new CoA at Association/Reassociation; for example, a DHCP server may assign a CoA not previously given to the mobile node. When dynamic VLAN assignment is used, the VLAN assignment is not even determined until IEEE 802.1X authentication has completed, which is after Association/Reassociation in [IEEE-802.11i].

In "Link Characteristics Information for Mobile IP" [Lee], link characteristics are included in registration/Binding Update messages sent by the mobile node to the home agent and correspondent node. Where the mobile node is acting as a receiver, this allows the correspondent node to adjust its transport parameters window more rapidly than might otherwise be possible. Link characteristics that may be communicated include the link type (e.g., 802.11b, CDMA (Code Division Multiple Access), GPRS (General Packet Radio Service), etc.) and link bandwidth. While the document suggests that the correspondent node should adjust its sending rate based on the advertised link bandwidth, this may not be wise in some circumstances. For example, where the mobile node link is not the bottleneck, adjusting the sending rate based on the link bandwidth could cause congestion. Also, where the transmission rate changes frequently, sending registration messages on each transmission rate

change could by itself consume significant bandwidth. Even where the advertised link characteristics indicate the need for a smaller congestion window, it may be non-trivial to adjust the sending rates of individual connections where there are multiple connections open between a mobile node and correspondent node. A more conservative approach would be to trigger parameter re-estimation and slow start based on the receipt of a registration message or Binding Update.

In "Hotspot Mitigation Protocol (HMP)" [HMP], it is noted that Mobile Ad-hoc NETwork (MANET) routing protocols have a tendency to concentrate traffic since they utilize shortest-path metrics and allow nodes to respond to route queries with cached routes. The authors propose that nodes participating in an ad-hoc wireless mesh monitor local conditions such as MAC delay, buffer consumption, and packet loss. Where congestion is detected, this is communicated to neighboring nodes via an IP option. In response to moderate congestion, nodes suppress route requests; where major congestion is detected, nodes rate control transport connections flowing through them. The authors argue that for ad-hoc networks, throttling by intermediate nodes is more effective than end-to-end congestion control mechanisms.

A.3. Transport Layer

Within the transport layer, proposals have focused on countering the effects of handoff-induced packet loss and non-congestive loss caused by lossy wireless links.

Where a mobile host moves to a new network, the transport parameters (including the RTT, RTO, and congestion window) may no longer be valid. Where the path change occurs on the sender (e.g., change in outgoing or incoming interface), the sender can reset its congestion window and parameter estimates. However, where it occurs on the receiver, the sender may not be aware of the path change.

In "The BU-trigger method for improving TCP performance over Mobile IPv6" [Kim], the authors note that handoff-related packet loss is interpreted as congestion by the transport layer. In the case where the correspondent node is sending to the mobile node, it is proposed that receipt of a Binding Update by the correspondent node be used as a signal to the transport layer to adjust cwnd and ssthresh values, which may have been reduced due to handoff-induced packet loss. The authors recommend that cwnd and ssthresh be recovered to pre-timeout values, regardless of whether the link parameters have changed. The paper does not discuss the behavior of a mobile node sending a Binding Update, in the case where the mobile node is sending to the correspondent node.

In "Effect of Vertical Handovers on Performance of TCP-Friendly Rate Control" [Gurtov], the authors examine the effect of explicit handover notifications on TCP-friendly rate control (TFRC). Where explicit handover notification includes information on the loss rate and throughput of the new link, this can be used to instantaneously change the transmission rate of the sender. The authors also found that resetting the TFRC receiver state after handover enabled parameter estimates to adjust more quickly.

In "Adapting End Host Congestion Control for Mobility" [Eddy], the authors note that while MIPv6 with route optimization allows a receiver to communicate a subnet change to the sender via a Binding Update, this is not available within MIPv4. To provide a communication vehicle that can be universally employed, the authors propose a TCP option that allows a connection endpoint to inform a peer of a subnet change. The document does not advocate utilization of "Link Up" or "Link Down" events since these events are not necessarily indicative of subnet change. On detection of subnet change, it is advocated that the congestion window be reset to INIT_WINDOW and that transport parameters be re-estimated. The authors argue that recovery from slow start results in higher throughput both when the subnet change results in lower bottleneck bandwidth as well as when bottleneck bandwidth increases.

In "Efficient Mobility Management for Vertical Handoff between WWAN and WLAN" [Vertical], the authors propose a "Virtual Connectivity Manager", which utilizes local connection translation (LCT) and a subscription/notification service supporting simultaneous movement in order to enable end-to-end mobility and maintain TCP throughput during vertical handovers.

In an early version of "Datagram Congestion Control Protocol (DCCP)" [RFC4340], a "Reset Congestion State" option was proposed in Section 11. This option was removed in part because the use conditions were not fully understood:

> An HC-Receiver sends the Reset Congestion State option to its sender to force the sender to reset its congestion state -- that is, to "slow start", as if the connection were beginning again.
> ...
> The Reset Congestion State option is reserved for the very few cases when an endpoint knows that the congestion properties of a path have changed. Currently, this reduces to mobility: a DCCP endpoint on a mobile host MUST send Reset Congestion State to its peer after the mobile host changes address or path.

"Framework and Requirements for TRIGTRAN" [TRIGTRAN] discusses optimizations to recover earlier from a retransmission timeout incurred during a period in which an interface or intervening link was down. "End-to-end, Implicit 'Link-Up' Notification" [E2ELinkup] describes methods by which a TCP implementation that has backed off its retransmission timer due to frame loss on a remote link can learn that the link has once again become operational. This enables retransmission to be attempted prior to expiration of the backed-off retransmission timer.

"Link-layer Triggers Protocol" [Yegin] describes transport issues arising from lack of host awareness of link conditions on downstream Access Points and routers. Transport of link layer triggers is proposed to address the issue.

"TCP Extensions for Immediate Retransmissions" [Eggert] describes how a transport layer implementation may utilize existing "end-to-end connectivity restored" indications. It is proposed that in addition to regularly scheduled retransmissions that retransmission be attempted by the transport layer on receipt of an indication that connectivity to a peer node may have been restored. End-to-end connectivity restoration indications include "Link Up", confirmation of first-hop router reachability, confirmation of Internet layer configuration, and receipt of other traffic from the peer.

In "Discriminating Congestion Losses from Wireless Losses Using Interarrival Times at the Receiver" [Biaz], the authors propose a scheme for differentiating congestive losses from wireless transmission losses based on inter-arrival times. Where the loss is due to wireless transmission rather than congestion, congestive backoff and cwnd adjustment is omitted. However, the scheme appears to assume equal spacing between packets, which is not realistic in an environment exhibiting link layer frame loss. The scheme is shown to function well only when the wireless link is the bottleneck, which is often the case with cellular networks, but not with IEEE 802.11 deployment scenarios such as home or hotspot use.

In "Improving Performance of TCP over Wireless Networks" [Bakshi], the authors focus on the performance of TCP over wireless networks with burst losses. The authors simulate performance of TCP Tahoe within ns-2, utilizing a two-state Markov model, representing "good" and "bad" states. Where the receiver is connected over a wireless link, the authors simulate the effect of an Explicit Bad State Notification (EBSN) sent by an Access Point unable to reach the receiver. In response to an EBSN, it is advocated that the existing retransmission timer be canceled and replaced by a new dynamically

estimated timeout, rather than being backed off. In the simulations, EBSN prevents unnecessary timeouts, decreasing RTT variance and improving throughput.

In "A Feedback-Based Scheme for Improving TCP Performance in Ad-Hoc Wireless Networks" [Chandran], the authors proposed an explicit Route Failure Notification (RFN), allowing the sender to stop its retransmission timers when the receiver becomes unreachable. On route reestablishment, a Route Reestablishment Notification (RRN) is sent, unfreezing the timer. Simulations indicate that the scheme significantly improves throughput and reduces unnecessary retransmissions.

In "Analysis of TCP Performance over Mobile Ad Hoc Networks" [Holland], the authors explore how explicit link failure notification (ELFN) can improve the performance of TCP in mobile ad hoc networks. ELFN informs the TCP sender about link and route failures so that it need not treat the ensuing packet loss as due to congestion. Using an ns-2 simulation of TCP Reno over 802.11 with routing provided by the Dynamic Source Routing (DSR) protocol, it is demonstrated that TCP performance falls considerably short of expected throughput based on the percentage of the time that the network is partitioned. A portion of the problem was attributed to the inability of the routing protocol to quickly recognize and purge stale routes, leading to excessive link failures; performance improved dramatically when route caching was turned off. Interactions between the route request and transport retransmission timers were also noted. Where the route request timer is too large, new routes cannot be supplied in time to prevent the transport timer from expiring, and where the route request timer is too small, network congestion may result.

For their implementation of ELFN, the authors piggybacked additional information (sender and receiver addresses and ports, the TCP sequence number) on an existing "route failure" notice to enable the sender to identify the affected connection. Where a TCP receives an ELFN, it disables the retransmission timer and enters "stand-by" mode, where packets are sent at periodic intervals to determine if the route has been reestablished. If an acknowledgment is received, then the retransmission timers are restored. Simulations show that performance is sensitive to the probe interval, with intervals of 30 seconds or greater giving worse performance than TCP Reno. The effect of resetting the congestion window and RTO values was also investigated. In the study, resetting the congestion window to one did not have much of an effect on throughput, since the bandwidth/delay of the network was only a few packets. However, resetting the RTO to a high initial value (6 seconds) did have a substantial detrimental effect, particularly at high speed. In terms of the probe packet sent, the simulations showed little difference

between sending the first packet in the congestion window, or retransmitting the packet with the lowest sequence number among those signaled as lost via the ELFNs.

In "Improving TCP Performance over Wireless Links" [Goel], the authors propose use of an ICMP-DEFER message, sent by a wireless Access Point on failure of a transmission attempt. After exhaustion of retransmission attempts, an ICMP-RETRANSMIT message is sent. On receipt of an ICMP-DEFER message, the expiry of the retransmission timer is postponed by the current RTO estimate. On receipt of an ICMP-RETRANSMIT message, the segment is retransmitted. On retransmission, the congestion window is not reduced; when coming out of fast recovery, the congestion window is reset to its value prior to fast retransmission and fast recovery. Using a two-state Markov model, simulated using ns-2, the authors show that the scheme improves throughput.

In "Explicit Transport Error Notification (ETEN) for Error-Prone Wireless and Satellite Networks" [Krishnan], the authors examine the use of explicit transport error notification (ETEN) to aid TCP in distinguishing congestive losses from those due to corruption. Both per-packet and cumulative ETEN mechanisms were simulated in ns-2, using both TCP Reno and TCP SACK over a wide range of bit error rates and traffic conditions. While per-packet ETEN mechanisms provided substantial gains in TCP goodput without congestion, where congestion was also present, the gains were not significant. Cumulative ETEN mechanisms did not perform as well in the study. The authors point out that ETEN faces significant deployment barriers since it can create new security vulnerabilities and requires implementations to obtain reliable information from the headers of corrupt packets.

In "Towards More Expressive Transport-Layer Interfaces" [Eggert2], the authors propose extensions to existing network/transport and transport/application interfaces to improve the performance of the transport layer in the face of changes in path characteristics varying more quickly than the round-trip time.

In "Protocol Enhancements for Intermittently Connected Hosts" [Schuetz], the authors note that intermittent connectivity can lead to poor performance and connectivity failures. To address these problems, the authors combine the use of the Host Identity Protocol (HIP) [RFC4423] with a TCP User Timeout Option and TCP Retransmission trigger, demonstrating significant improvement.

A.4. Application Layer

In "Application-oriented Link Adaptation for IEEE 802.11" [Haratcherev2], rate information generated by a link layer utilizing improved rate adaptation algorithms is provided to a video application, and used for codec adaptation. Coupling the link and application layers results in major improvements in the Peak Signal to Noise Ratio (PSNR). Since this approach assumes that the link represents the path bottleneck bandwidth, it is not universally applicable to use over the Internet.

At the application layer, the usage of "Link Down" indications has been proposed to augment presence systems. In such systems, client devices periodically refresh their presence state using application layer protocols such as SIP for Instant Messaging and Presence Leveraging Extensions (SIMPLE) [RFC3428] or Extensible Messaging and Presence Protocol (XMPP) [RFC3921]. If the client should become disconnected, their unavailability will not be detected until the presence status times out, which can take many minutes. However, if a link goes down, and a disconnect indication can be sent to the presence server (presumably by the Access Point, which remains connected), the status of the user’s communication application can be updated nearly instantaneously.

Appendix B. IAB Members at the Time of This Writing

Bernard Aboba
Loa Andersson
Brian Carpenter
Leslie Daigle
Elwyn Davies
Kevin Fall
Olaf Kolkman
Kurtis Lindqvist
David Meyer
David Oran
Eric Rescorla
Dave Thaler
Lixia Zhang

Author's Address

Bernard Aboba, Ed.
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

EMail: bernarda@microsoft.com
Phone: +1 425 706 6605
Fax: +1 425 936 7329

IAB

EMail: iab@iab.org
URI: http://www.iab.org/

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.