Network Working Group B. Aboba, Ed.
INTERNET-DRAFT Internet Architecture Board
Category: Informational IAB
<draft-iab-link-indications-03.txt>
18 August 2005

Architectural Implications of Link Indications

Abstract

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 layer 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.

This document provides an overview of the role of link indications within the Internet Architecture. While the judicious use of link indications can provide performance benefits, experience has also shown that that inappropriate use can degrade both robustness and performance.

This document summarizes the current understanding of the role of link indications, and provides advice to document authors about the appropriate use of link indications.

In 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 presents a summary of the literature on link indication utilization.

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.

Association
The service used to establish an access point/station (AP/STA) mapping and enable stations to access the Distribution System network via the wireless medium.

Basic Service Set (BSS)
An IEEE 802.11 specific term. A set of stations controlled by a single coordination function, where the coordination function may be centralized (e.g., in a single AP) or distributed (e.g., for an ad-hoc network). Membership of a BSS does not imply that wireless communication with all other members of the BSS is possible.

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.

Care of Address (CoA)
A unicast routable address associated with a mobile node while visiting a foreign link; the subnet prefix of this IP address is a foreign subnet prefix. Among the multiple care-of addresses that a mobile node may have at any given time (e.g., with different subnet prefixes), the one registered with the mobile node's home agent for a given home address is called its "primary" care-of address.

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

Distribution System (DS)
A system used to interconnect a set of basic service sets (BSSs) and integrated local area networks (LANs) to create an extended service set (ESS).

Dynamic Host Configuration Protocol (DHCP) client
A DHCP client is an Internet host using DHCP to obtain configuration parameters such as a network address.

DHCP server
A DHCP server or "server" is an Internet host that returns configuration parameters to DHCP clients.

Distribution System (DS)
A system used to interconnect a set of basic service sets (BSSs) and integrated local area networks (LANs) to create an extended service set (ESS). The Distribution System is a network connecting Access Points, thereby enabling wider wireless coverage than a single access point can provide.

Extended Service Set (ESS)
A set of one or more interconnected basic service sets (BSSs) that appears as a single BSS to the logical link control layer at any station associated with one of those BSSs. TheWhile link indications may show promise, it may be difficult to prove that processing of a given indication provides benefits in a wide

variety of circumstances. ESS can be thought of as the coverage area provided by a collection of APs all interconnected by the Distribution System. It may consist of one or more prefixes.

Independent Basic Service Set (IBSS)
A BSS that forms a self-contained network, and in which no access to a distribution system (DS) is available.

Inter-Access Point Protocol (IAPP)
A protocol used between access points that assures that the station may only be connected to a single AP within the ESS at a time, and also provides for transfer of context to the new AP.

Link A communication facility or physical medium that can sustain data communications between multiple network nodes, such as an Ethernet (simple or bridged). A link is the layer immediately below IP. In a layered network stack model, the Link Layer (Layer 2) is normally below the Network (IP) Layer (Layer 3), and above the Physical Layer (Layer 1). Each link is associated with a minimum of two endpoints. Each link endpoint has a unique link-layer identifier.

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

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

Link identifier
An indication provided by the link layer as to which network(s) a host has connected to. Examples include the SSID with IEEE 802.11. For details, see [DNAv4] Appendix A.

Link indication
Information provided by the link layer to higher layers regarding the state of the link. In addition to "Link Up" and "Link Down",

relevant information may include the current link rate, link identifiers (e.g. SSID, BSSID in 802.11), and link performance statistics (such as the delay or loss rate).

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

Most Likely Networks (MLNs)
The attached network(s) heuristically determined by the host to be most likely.

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

Medium Access Protocol (MAC)
A protocol for mediating access to, and possibly allocation of, the physical communications medium. Nodes participating in the medium access protocol can communicate only when they have uncontested access to the medium, so that there will be no interference. When the physical medium is a radio channel, the MAC is the same as the Channel Access Protocol.

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
The term "operable address" refers to either a static address, or a dynamically assigned address which has not been relinquished, and has not expired.

Routable address
In this specification, the term "routable address" refers to any address other than an IPv4 Link-Local address [RFC3927]. This includes private addresses as specified in [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).

Service Set Identifier (SSID)
The SSID indicates the identity of an ESS or IBSS.

Weak End-System Model
In the Weak End-System Model, packets sent out an interface need

not necessarily have a source address configured on that interface.

1.3. Overview

Link status was first taken into account in computer routing within the ARPANET as early as 1969. In response to an attempt to send to a host that was off-line, the ARPANET link layer protocol provided a "Destination Dead" indication [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 [ETX].

"Routing Information Protocol" [RFC1058] defines RIP, which is descended from the Xerox Network Systems (XNS) Routing Information Protocol. "The Open Shortest Path First Specification" [RFC1131] defines OSPF, which uses Link State Advertisements (LSAs) in order to flood information relating to link status within an OSPF area. 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.

In 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 or frame loss due to interference or signal fading. In both wired and wireless links, the link state may rapidly flap between the "up" and "down" states.

Routing protocol implementations have had to take real-world wired link behavior into account in order to maintain robustness. 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.

"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 bi-directional reachability has been confirmed.

More recently, the importance of realistic wireless link models has become better appreciated. 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, [Kotz] notes 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, signal and signal/noise ratio.

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 rates), 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.

The complexity of real-world link behavior poses a challenge to the integration 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 provides guidance on the incorporation of link indications within the Internet, Transport and Application layers.

1.4. Layered Indication Model

A layered indication model is shown in Figure 1 which includes both internally generated link indications and indications arising from external interactions such as receipt of Mobile IP Binding Updates, and path change detection.

In this model, link indications include frame loss (before retransmissions), the current link rate, the link state (up/down), and link identifiers. These indications may be inter-dependent, since rate adjustment and detection algorithms are typically influenced by frame loss, and a "Link Down" indication may be influenced by the detection and search process. Link identifiers are typically obtained in the process of bringing the link up.

1.4.1. Internet Layer

The Internet layer is the primary user of link indications, since one of its functions is to shield applications from the specifics of link behavior. The Internet layer utilizes link indications in order to to optimize aspects of IP configuration, routing, and mobility. By validating and filtering link indications and selecting outgoing and incoming interfaces based on routing metrics, the Internet layer enables upper layers to avoid dependency on link indications.

In "Detecting Network Attachment" [DNAv4], "Link Up" indications and link identifiers are used as hints for validating an existing IP configuration. Once the IP configuration is confirmed, it may be determined that an address change has occurred. However, "Link Up" indications often do not result in a change to Internet layer configuration.

The routing sub-layer utilizes link indications in order to calculate routing metrics and determine changes in link state. As described in [Iannaccone], damping of link flaps and rate limiting of link state advertisements are examples of how the routing sub-layer validates and filters link indications.

Routing metrics incorporating link layer indications enable gateways to obtain knowledge of path changes and take remote link conditions into account for the purposes of route selection. When a link experiences frame loss, routing metrics incorporating frame loss such as the metrics described in [ETX][ETX-Rate][ETX-Radio] increase, possibly resulting in selection of an alternate route. If a troubled link represents the only path to a prefix and the link experiences high frame loss ("down"), the route will be withdrawn or the metric will become infinite. Similarly, when the link becomes operational, the route will appear again. Where routing protocol security is

implemented, this information can be securely propagated.

Within "Weak End-System Model" 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 link 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 by running a routing protocol, receipt of a Router Advertisement, dead gateway detection [RFC816] or a change in the IP TTL 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).

1.4.2. Transport Layer

The Transport layer processes Internet layer and link indications differently for the purposes of transport parameter estimation and connection management. For the purposes of parameter estimation, the Transport layer may be interested in a wide range of Internet and link layer indications. The Transport layer may wish to use path change indications from the Internet layer in order to reset parameter estimates. It may also be useful for the Transport layer to utilize link layer indications such as link rate, frame loss rate and "Link Up"/"Link Down" in order to improve transport parameter estimates.

As described in Section A.3, the algorithms for improving transport parameter estimates using link layer indications are still under development. In transport parameter estimation, layering considerations do not exist to the same extent as in connection management. For example, the Internet layer may receive a "Link Down" indication followed by a subsequent "Link Up" indication. This information may be useful for transport parameter estimation even if IP configuration does not change, since it may indicate the potential for non-congestive packet loss during the period between the indications.

For the purposes of connection management, the Transport layer typically only utilizes Internet layer indications such as changes in the incoming/outgoing interface and IP configuration changes. For example, the Transport layer may tear down transport connections due to invalidation of a connection endpoint IP address. However, before this can occur, the Internet layer must determine that a

configuration change has occurred.

Nevertheless, the Transport layer does not respond to all Internet layer indications. For example, an Internet layer configuration change may not be relevant for the purposes of connection management. Where the connection has been established based on the 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.

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 PPP LCP-Terminate or an 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. "Requirements for Internet Hosts - Communication Layers" [RFC1122] [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] 4.2.3.9 requires TCP to react to an 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 "Path MTU Discovery" [RFC1191] use Destination Unreachable code 4, they do not treat this as a hard error condition.

However, "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.

1.4.3. Application Layer

In addition to Internet layer indications propagated to the Application layer (such as IP address configuration and changes), the Transport layer provides its own indications to the Application layer, such as connection teardown. The Transport layer may also provide indications to the link layer. For example, to prevent excessive retransmissions within the link layer, where the link layer retransmission timeout is significantly less than the path round-trip timeout, 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.

```
                 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   Application   |                                               |
   Layer         |                                               |
                 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                                                     ^   ^
                                                     !   !
                 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-!-+-!-+-+-+-+
                 |                                   !   !       |
                 |                                   ^   ^       |
                 |     Connection Management         ! Teardown  |
   Transport     |                                   !           |
   Layer         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-!-+-+-+-+-+-+
                 |                                   !           |
                 | Transport Parameter Estimation    !           |
                 | Estimation (MTU, RTT, RTO, cwnd,  ! ssthresh)
                 |                                   ^           |
                 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-!-+-+-+-+-+-+
                       ^           ^       ^         !
                       !           !       !         !
                 +-+-+-!-+-+-+-+-+-!-+-+-+-!-+-+-+-+-!-+-+-+-+-+-+
                 |     ! Incoming  !MIP    !         !           |
                 |     ! Interface !BU     !         !           |
                 |     ! Change    !Receipt!         !           |
                 |     ^           ^       ^         ^           |
   Internet      |     ! Mobility  !       !         !           |
   Layer         +-+-+-!-+-+-+-+-+-!-+-+-+-!-+-+-+-+-!-+-+-+-+-+-+
                 |     ! Outgoing  ! Path  !         !           |
                 |     ! Interface ! Change!         !           |
                 |     ^ Change    ^       ^         ^           |
                 |                         !         !           |
                 |       Routing           !         !           |
                 | ^         ^             !         !           |
                 +-!-+-+-+-+-!-+-+-+-+-+-+-!-+-+-+-+-!-+-+-+-+-+-+
                 | !         !             !         ! IP        |
                 | !         !             !         ! Address   |
                 | !     IP  !Configuration^         ^ Config/   |
                 | !         !             !           Changes   |
                 +-!-+-+-+-+-!-+-+-+-+-+-+-!-+-+-+-+-+-+-+-+-+-+-+
                   !         !             !       ^
                   !         !             !       !
                 +-!-+-+-+-+-!-+-+-+-+-+-+-!-+-+-+-!-+-+-+-+-+-+-+
                 | !         !             !       !             |
   Link          | ^         ^             ^       ^             |
   Layer         | Frame    Rate          Link     Link          |
                 | Loss     Adjustment   Up/Down  Identifiers    |
                 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1. Layered Indication Model

In 802.11, this can be achieved by adjusting the MIB 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. Also, 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.

Since applications can obtain the information they need from Internet and Transport layer indications they should not 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. As a result, applications should utilize an Internet layer "IP Address Configured" event instead of a "Link Up" indication. Similarly, applications should not utilize "Link Down" indications, since they can be rapidly followed by a "Link Up" indication; instead, they should respond to Transport layer teardown indications.

2. Architectural Considerations

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 misleading 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 advocating the transport of link indications beyond the local host need to carefully consider the layering, security and transport implications (Section 2.9). In general, implicit signals are preferred to explicit transport of link indications since they 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.

2.1. Model Validation

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

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 dependent on link indications should explicitly articulate the assumptions of the link model and describe the circumstances in which it applies.

For example, generic "trigger" models often include implicit assumptions which may prove invalid in outdoor or mesh deployments. For example, two-state Markov models where the link is either in a state experiencing low frame loss ("up") or in a state where few frames are successfully delivered ("down") have frequently been used. 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.

Link indications based on signal quality "Link Quality Crosses Threshold" typically assume the absence of multi-path interference, so that signal to noise ratio varies smoothly in space, and frame loss is well predicted by signal strength and distance.

However, where multi-path interference is present, signal strength and signal/noise ratio can vary rapidly and high signal/noise ratio can co-exist with high frame loss. Where links may exist in intermediate states between "up" and "down" or asymmetry is encountered, a "Link Quality Crosses Threshold" indication may exhibit excessive jitter and may prove to be unreliable predictors of future link performance.

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.

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.

[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] 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".

Link indication definitions should head 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 networks.

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 effects such as multi-path interference, signal strength and signal/noise ratio may vary rapidly over a short distance, causing rapid variations in frame loss and rate, and jitter in link indications based on these metrics. This can create problems for upper layers that act on these indications without sufficient damping.

[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 either generated 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, that it will 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 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.

2.3. Robustness

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

a. Implementation Variation
b. Recovery from invalid indications
c. 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.

As described in [Aguayo], wireless links often exhibit loss rates intermediate between "up" (low loss) and "down" (high loss) states, as well as substantial asymmetry. In these circumstances, a "Link Up" indication may not imply bi-directional reachability. Also, a reachability demonstration based on small packets may not mean that the link is suitable for carrying larger data packets. As a result, "Link Up" and "Link Down" indications may not reliably determine whether a link is suitable for carrying IP data packets.

Where multi-path interference or hidden nodes are encountered, frame loss may vary widely over a short distance. While techniques such as use of multiple antennas may be used to reduce multi-path effects and RTS/CTS signaling can be used to address hidden node problems, these techniques may not be completely effective. As a result, a mobile host may find itself experiencing widely varying link conditions, causing the link to rapidly cycle between "up" and "down" states, with "Going down" or "Going up" indications providing little predictive value.

Where the reliability of a link layer indication is suspect, it is best for upper layers to treat the indication as a "hint" (advisory in nature), rather than a "trigger" forcing a given action. In order to provide increased robustness, heuristics can be developed to assist upper layers in determining whether the "hint" is valid or should be discarded.

To provide robustness in the face of potentially misleading link indications, in [DNAv4] "Link Up" indications are assumed to be inherently unreliable, so that bi-directional reachability needs to

be demonstrated in the process of validating an existing IPv4 configuration. However, where a link exhibits an intermediate loss rate, the success of the [DNAv4] reachability test does not guarantee that the link is suitable for carrying IP data packets.

Another example of link indication 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. 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.

Recovery is supported within [DNAv4] in the case where link indications may lead a host to erroneously conclude that the link prefix remains unchanged when the host has in fact changed networks. In this case, the bi-directional reachability test times out, and the host will eventually realize its mistake and obtain an IP address by normal means.

Where a proposal involves recovery at the transport layer, the recovered transport parameters (such as the MTU, RTT, RTO, congestion window, etc.) must be demonstrated to remain valid. Congestion window validation is discussed in [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 performance, including changes in the negotiated rate, 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 many 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 a IPv4 Link-Local address prior to determining whether a DHCP server is available [RFC3927].

As noted in Section 1.4, the Transport layer does not utilize "Link Up" and "Link Down" indications for the purposes of connection management. Since applications can obtain the information they need from Internet and Transport layer indications they should not utilize link indications.

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:

[a] Rate limiting. Packets generated by the receipt of link indications can be rate limited (e.g. a limit of one packet per end-to-end path RTO).

[b] 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".

[c] Keepalives. Instead of utilizing a "Link Down" indication, an application can utilize an application keepalive or Transport layer indication such as connection teardown.

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

Note that congestion control is not solely an issue for the transport layer, nor is "conservation of packets" sufficient to avoid congestive collapse in all cases. Link layer algorithms that adjust rate based on frame loss also need to demonstrate conservatism in the face of congestion. For example, "Roaming Interval Measurements" [Alimian] demonstrates that 802.11 implementations show wide variation in rate adaptation behavior. This is worrisome, since implementations that rapidly decrease the negotiated rate in response to frame loss can cause congestive collapse in the link layer, even where exponential backoff is implemented. For example, an implementation that decreases rate by a factor of two while backing off the retransmission timer by a factor of two has not reduced consumption of available slots within the MAC. While such an implementation might demonstrate "conservation of packets" it does not conserve critical resources.

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" when link indications are 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 have an operable Internet layer configuration.

2.5. Effectiveness

Proposals must demonstrate the effectiveness of proposed optimizations. It may be difficult to prove that a given indication provides benefits in a wide variety of circumstances. Since optimizations often carry a burden of increased complexity, substantial performance improvement is required to make a compelling case.

In the face of unreliable link indications, effectiveness may depend heavily on the penalty for false positives and false negatives. As noted in [DNAv4], it is simultaneously possible for a link indication to be highly reliable and provide no net benefit, depending on the probability of a false indication and the penalty paid for the false indication. In the case of [DNAv4], the benefits of successful optimization are modest, but the penalty for falsely concluding that the network remains unchanged is a lengthy timeout. The result is that link indications may not be worth considering if they are incorrect more than a small fraction of the time.

For example, it can be argued that a change in the Service Set Identifier (SSID) in [IEEE-802.11] is not a sufficiently reliable indication of a prefix change. Within IEEE 802.11, the Service Set Identifier (SSID) functions as a non-unique identifier of the administrative domain of a Wireless LAN. Since the SSID is non-unique, many different operators may share the same SSID, and Access Points typically ship with a default value for the SSID (e.g. "default"). Since the SSID relates to the administrative domain and not the network topology, multiple SSIDs may provide access to the same prefix, and a single SSID may provide access to multiple prefixes at one or multiple locations.

Given this, it is unreliable to use the SSID alone for the purpose of movement detection. A host moving from one point of attachment to another, both with the same SSID, may have remained within the same network, or may have changed networks. Similarly, a host discovering that the SSID has changed may have changed networks, or it may not have. Moreover, where private address space is in use, it is possible for the SSID, the prefix (e.g. 192.168/16) and even the default gateway IP address to remain unchanged, yet for the host to have moved to a different network. Were the host to make decisions relating to configuration of the IP layer (such as address

assignment) based solely on the SSID, address conflicts are likely.

2.6. Interoperability

Link indications should not be required by upper layers, in order to maintain link independence.

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

For example, if link layer prefix hints are provided as a substitute for Internet layer configuration, hosts not understanding those hints would be unable 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).

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 [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 on one of its interfaces, a routing metric taking frame loss 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 for the original interface (MTU,

congestion state) may no longer be valid for the new outgoing and incoming paths.

To avoid race conditions, the following measures are recommended:

a. Path change processing
b. Layering
c. Metric consistency

2.7.1. Path Change Processing

When the Internet layer detects a path change, such as a change in the outgoing or incoming interface of the host or the incoming interface of a peer, or perhaps a substantial change in the TTL of received IP packets, it may be worth considering whether to reset transport parameters (RTT, RTO, cwnd, MTU) to their initial values and 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. Since one of the major objectives of the Internet layer is maintaining link layer independence, upper layers relying on Internet layer indications rather than consuming link indications directly can avoid link layer dependencies.

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. Once a link is in the "up" state, its effectiveness in transmission of data packets can be determined. For example, frame loss may be used to assist in rate adjustment and to determine when to select an alternative point of attachment. Also, the effective throughput depends on the negotiated rate and frame loss, and can be used in calculation of the routing metric, as described in [ETX][ETX-Rate][ETX-Radio].

However, prior to sending data packets over the link, other metrics are required to determine suitability. 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 data packets.

Since the negotiated rate and frame loss typically cannot be predicted prior to utilizing the link for data traffic, existing implementations often utilize metrics such as signal strength and access point load in handoff decisions. The "Link Going Down", "Link Going Up", "Link Quality Crosses Threshold" indications were developed primarily to assist with handoff between interfaces, and are oriented toward inferred rather than measured suitability.

Research indicates that this approach may have some promise. In order to enable stations to roam prior to encountering packet loss, studies such as [Vatn] have suggested using signal strength as a detection mechanism, rather than frame loss, as suggested in [Velayos]. [Vertical] proposes use of signal strength and link utilization in order to optimize vertical handoff and demonstrates improved TCP throughput.

However, 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. For example, receipt of "Link Going Down" or "Link Quality Crosses Threshold" indications could be used as a signal to enable another interface. 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.

As noted in [Aguayo], signal strength and distance are not good predictors of frame loss or negotiated rate, 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 negotiated rate. Similarly, an AP demonstrating low utilization may not necessarily be the best choice, since utilization may be low due to hardware or software problems. [Villamizar] notes that link utilization-based routing metrics have a history of instability, so that they are rarely deployed.

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 re-establishment; 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 EAP 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 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 including the transport of link indications need to carefully consider the layering, security and transport implications. In general, implicit signals are preferred to explicit transport of link indications since they 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.

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

[a] Absence of alternatives. By default, alternatives not requiring explicit signaling are preferred. Where these solutions are shown to be inadequate, proposals must prove that existing explicit signaling mechanisms (such as path change processing and link-aware routing metrics) are inadequate.

[b] Mitigation of security issues. Proposals need to describe how security issues can be addressed. A host receiving a link indication from a router typically will not be able to authenticate the indication. Where indications can be transported over the Internet, this allows an attack to be launched without requiring access to the link.

[c] Validation of transported indications. Even if a transported link indication can be 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 saying something about Internet, Transport or Application layer operations at a remote element. These layers use different identifiers, and so it is necessary to match the link indication with relevant higher layer state. Therefore proposals need to demonstrate how the link indication can be mapped to the right higher layer state. For example, if a presence server is receiving remote indications of "Link Up"/"Link Down" status for a particular 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

While Figure 1 presents an overview of how link indications are utilized by the Internet, Transport and Application layers, further work is needed in this area.

At the Link and Internet layers, more work is needed to reconcile pre and post-connection metrics, such as reconciling metrics utilized in handoff (e.g. signal strength and link utilization) with link-aware routing metrics (e.g. frame loss and negotiated rate).

More work is also needed in the area of link-aware routing metrics. Since [IEEE-802.11e] incorporates burst ACKs, the relationship between 802.11 link throughput and frame loss is growing more complex. This may necessitate the development of revised routing

metrics, taking the more complex retransmission behavior into account. More work is also needed in order to apply link-aware routing metrics to host behavior.

At the Transport layer, more work is needed to determine the appropriate reaction to Internet layer indications such as path changes. For example, it may make sense for the Transport layer to adjust transport parameter estimates in response to "Link Up"/"Link Down" indications and frame loss, so that transport parameters are not adjusted as though congestion were detected when loss is occurring in the link layer or a "Link Down" indication has been received.

Finally, more work is 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.

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 802.11 does not protect management frames such as Associate/ Reasociate, 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 be 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, 802.11 wireless LANs implementing [IEEE-802.11i] do not enable stations to send or receiving IP packets on the link until completion of an authenticated key exchange protocol known as the "4-way handshake". As a result, an 802.11 link utilizing [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 (AP).

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 gateways 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 gateways 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 be 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 particular careful to avoid enabling denial of service attacks that can 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, greatly expand the level of vulnerability.

By enabling the transport of link indications, it is possible to transform an attack that might otherwise be restricted to attackers on the local link into one which can 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, either by generating multiple responses or a single larger response. For example, link indication proposals should not enable multiple hosts to respond to a frame with a multicast destination address.

5. References

5.1. Informative References

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

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

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

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

[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", RFC 1918, February 1996.

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

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

[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.

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

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

[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.

[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", draft-ietf-bfd-base-02.txt, Internet draft (work in progress), March 2005.

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

[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.

[DCCP] Kohler, E., Handley, M. and S. Floyd, "Datagram Congestion Control Protocol (DCCP)", Internet drafts (work in progress), draft-ietf-dccp-spec-08.txt, October 2004.

[DNAv4] Aboba, B., "Detection of Network Attachment in IPv4", draft-ietf-dhc-dna-ipv4-14.txt, Internet draft (work in progress), August 2005.

[E2ELinkup] Dawkins, S. and C. Williams, "End-to-end, Implicit 'Link-Up' Notification", draft-dawkins-trigtran-linkup-01.txt, Internet draft (work in progress), October 2003.

[EAPIKEv2] Tschofenig, H., D. Kroeselberg and Y. Ohba, "EAP IKEv2 Method", draft-tschofenig-eap-ikev2-05.txt, Internet draft (work in progress), October 2004.

[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.

[Eggert] Eggert, L., Schuetz, S. and S. Schmid, "TCP Extensions for Immediate Retransmissions", draft-eggert-tcpm-tcp-retransmit-now-01.txt, Internet draft (work in progress), September 2004.

[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", draft-gont-tcpm-icmp-attacks-03.txt, Internet draft (work in progress), December 2004.

[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).

[HMP] Lee, S., Cho, J. and A. Campbell, "Hotspot Mitigation Protocol (HMP)", draft-lee-hmp-00.txt, Internet draft (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, "Draft Amendment 7: Medium Access Control (MAC) Quality of Service (QoS) Enhancements", IEEE 802.11e Draft 10.0, October 2004.

[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.

[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/D2.0, February 2005.

[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.

[Kim] Kim, K., Park, Y., Suh, K., and Y. Park, "The BU-trigger method for improving TCP performance over Mobile IPv6", draft-kim-tsvwg-butrigger-00.txt, Internet draft (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., Allman, M., Partridge, P. and J. Sterbenz, "Explicit Transport Error Notification (ETEN) for Error-Prone Wireless and Satellite Networks", Technical Report No. 8333, BBN Technologies, March 2002.

[Lee] Park, S., Lee, M. and J. Korhonen, "Link Characteristics Information for Mobile IP", draft-daniel-mip-link-characteristic-01.txt, Internet draft (work in progress), April 2005.

[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", draft-giaretta-mip6-authorization-eap-02.txt, Internet draft (work in progress), October 2004.

[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.

[Mitani] Mitani, K., Shibui, R., Gogo, K. and F. Teraoka, "Unified L2 Abstractions for L3-Driven Fast Handover", draft-koki-mobopts-l2-abstractions-02.txt, Internet draft (work in progress), February 2005.

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

[Mun] Mun, Y. and J. Park, "Layer 2 Handoff for Mobile-IPv4 with 802.11", draft-mun-mobileip-layer2-handoff-mipv4-01.txt, Internet draft (work in progress), March 2004.

[PEAP] Palekar, A., Simon, D., Salowey, J., Zhou, H., Zorn, G. and S. Josefsson, "Protected EAP Protocol (PEAP) Version 2", draft-josefsson-pppext-eap-tls-eap-10.txt, Internet draft (work in progress), October 2004.

[Park] Park, S., Njedjou, E. and N. Montavont, "L2 Triggers Optimized Mobile IPv6 Vertical Handover: The 802.11/GPRS Example", draft-daniel-mip6-optimized-vertical-handover-00.txt, July 2004. ,IP [PRNET] Jubin, J. and J. Tornow, "The DARPA packet radio network protocols", Proceedings of the IEEE, 75(1), January 1987.

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

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

[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.

[Swami] Swami, Y., Le, K., Eddy, W., "Lightweight Mobility Detection and Response (LMDR) Algorithm for TCP", draft-swami-tcp-lmdr-05, Internet draft (work in progress), February 2005.

[TRIGTRAN] Dawkins, S., Williams, C. and A. Yegin, "Framework and Requirements for TRIGTRAN", draft-dawkins-trigtran-framework-00.txt, Internet draft (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.

[Yegin] Yegin, A., "Link-layer Triggers Protocol", draft-yegin-l2-triggers-00.txt, Internet Draft (work in progress), June 2002.

[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)", draft-ietf-ospf-omp-02.txt, Internet draft (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.

Appendix A - Literature Review

This Appendix summarizes the literature on utilization of link indications within the Link, Internet, Transport and Application layers.

A.1 Link Layer

The characteristics of wireless links have been found to vary considerably depending on the environment. 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 either be assumed to operate very well or not at all.

"Link-level Measurements from an 802.11b Mesh Network" [Aguayo] analyzes 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 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 phases (connection to an alternative point of attachment). 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, 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 S/N ratio. 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 STAs implementations communicating with two APs. The study found that there is 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 S/N 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 measure 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 NTP, enabling a driver implementing SyncScan to work with legacy APs without requiring implementation of new protocols. The authors measure the distribution of inter-arrival times for stations implementing SyncScan, with excellent results.

"Roaming Interval Measurements" [Alimian] presents data on 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 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 continue 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 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 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 mobile 802.11 STA encounters a series of consecutive non-

acknowledged frames, 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, after three consecutive non-acknowledged frames. Only when no alternative 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 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 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.

In "Link-layer Enhancements for TCP/IP over GSM" [Ludwig], the authors describe how the 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 SCTP 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 backoff 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 slowstart 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 rereception of single or multiple packets.

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 one hundred 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.

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 "Detection of Network Attachment (DNA) in IPv4" [DNAv4], link indications are utilized to enable a host that has moved to a new point of attachment to rapidly confirm a currently operable configuration, rather than utilizing the DHCP protocol [RFC2131].

"A High-Throughput Path Metric for Multi-Hop Wireless Routing" [ETX] describes how routing metrics can be improved by taking link layer frame loss rates into account, enabling the selection of routes maximizing available throughput. While the proposed routing metric utilizes the Expected Transmission Count (ETX), it does not take the negotiated rate into account. 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 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. 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 data rate.

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 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, such as is described in [DNAv4].

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 the new point of attachment represents a change of network or not. 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, GPRS, 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 in congestion. Also, where link rates change frequently, sending registration messages on each 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 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 adhoc wireless mesh monitor local conditions such as MAC delay, buffer consumption and packets 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 throttle TCP connections flowing through them. The authors argue that for adhoc 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. 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 "Lightweight Mobility Detection and Response (LMDR) Algorithm for TCP" [Swami], 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 reestimated. 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 an early draft of [DCCP], a "Reset Congestion State" option was proposed in Section 4. This option was removed in part because the use conditions were not fully understood:

An Half-Connection 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 interarrival 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 on an existing "route failure" notice (sender and receiver addresses and ports, the TCP sequence number) 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 acknowledgement 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 affect of resetting the congestion window and RTO values was also

investigated. In the study, resetting 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 signalled 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" [Krishan], 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.

A.4 Application Layer

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 SIMPLE [RFC3428] or 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

Intellectual Property Statement

Acknowledgment

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