ï»¿

           Updates for the Back-to-Back Frame Benchmark in RFC 2544

Abstract

   Fundamental benchmarking methodologies for network interconnect
   devices of interest to the IETF are defined in RFC 2544. This memo
   updates the procedures of the test to measure the Back-to-Back Frames
   benchmark of RFC 2544, based on further experience.

   This memo updates Section 26.4 of RFC 2544.

Status of This Memo

   This document is not an Internet Standards Track specification; it is
   published for informational purposes.

   This document is a product of the Internet Engineering Task Force
   (IETF). It represents the consensus of the IETF community. It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG). Not all documents
   approved by the IESG are candidates for any level of Internet
   Standard; see Section 2 of RFC 7841.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   https://www.rfc-editor.org/info/rfc9004.

Table of Contents

1. Introduction

   The IETF's fundamental benchmarking methodologies are defined in
   [RFC2544], supported by the terms and definitions in [RFC1242].
   [RFC2544] actually obsoletes an earlier specification, [RFC1944].
   Over time, the benchmarking community has updated [RFC2544] several
   times, including the Device Reset benchmark [RFC6201] and the
   important Applicability Statement [RFC6815] concerning use outside
   the Isolated Test Environment (ITE) required for accurate
   benchmarking. Other specifications implicitly update [RFC2544], such
   as the IPv6 benchmarking methodologies in [RFC5180].

   Recent testing experience with the Back-to-Back Frame test and
   benchmark in Section 26.4 of [RFC2544] indicates that an update is
   warranted [OPNFV-2017] [VSPERF-b2b]. In particular, analysis of the
   results indicates that buffer size matters when compensating for
   interruptions of software-packet processing, and this finding
   increases the importance of the Back-to-Back Frame characterization
   described here. This memo provides additional rationale and the
   updated method.

   [RFC2544] provides its own requirements language consistent with
   [RFC2119], since [RFC1944] (which it obsoletes) predates [RFC2119].
   All three memos share common authorship. Today, [RFC8174] clarifies
   the usage of requirements language, so the requirements language
   presented in this memo are expressed in accordance with [RFC8174].
   They are intended for those performing/reporting laboratory tests to
   improve clarity and repeatability, and for those designing devices
   that facilitate these tests.

2. Requirements Language

   The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
   "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and
   "OPTIONAL" in this document are to be interpreted as described in
   BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all
   capitals, as shown here.

3. Scope and Goals

   The scope of this memo is to define an updated method to
   unambiguously perform tests, measure the benchmark(s), and report the
   results for Back-to-Back Frames (as described in Section 26.4 of
   [RFC2544]).

   The goal is to provide more efficient test procedures where possible
   and expand reporting with additional interpretation of the results.
   The tests described in this memo address the cases in which the
   maximum frame rate of a single ingress port cannot be transferred to
   an egress port without loss (for some frame sizes of interest).

   Benchmarks as described in [RFC2544] rely on test conditions with
   constant frame sizes, with the goal of understanding what network-
   device capability has been tested. Tests with the smallest size
   stress the header-processing capacity, and tests with the largest
   size stress the overall bit-processing capacity. Tests with sizes in
   between may determine the transition between these two capacities.
   However, conditions simultaneously sending a mixture of Internet
   (IMIX) frame sizes, such as those described in [RFC6985], MUST NOT be
   used in Back-to-Back Frame testing.

   Section 3 of [RFC8239] describes buffer-size testing for physical
   networking devices in a data center. Those methods measure buffer
   latency directly with traffic on multiple ingress ports that overload
   an egress port on the Device Under Test (DUT) and are not subject to
   the revised calculations presented in this memo. Likewise, the
   methods of [RFC8239] SHOULD be used for test cases where the egress-
   port buffer is the known point of overload.

4. Motivation

   Section 3.1 of [RFC1242] describes the rationale for the Back-to-Back
   Frames benchmark. To summarize, there are several reasons that
   devices on a network produce bursts of frames at the minimum allowed
   spacing; and it is, therefore, worthwhile to understand the DUT limit
   on the length of such bursts in practice. The same document also
   states:

   | Tests of this parameter are intended to determine the extent of
   | data buffering in the device.

   Since this test was defined, there have been occasional discussions
   of the stability and repeatability of the results, both over time and
   across labs. Fortunately, the Open Platform for Network Function
   Virtualization (OPNFV) project on Virtual Switch Performance (VSPERF)
   Continuous Integration (CI) [VSPERF-CI] testing routinely repeats
   Back-to-Back Frame tests to verify that test functionality has been
   maintained through development of the test-control programs. These
   tests were used as a basis to evaluate stability and repeatability,
   even across lab setups when the test platform was migrated to new DUT
   hardware at the end of 2016.

   When the VSPERF CI results were examined [VSPERF-b2b], several
   aspects of the results were considered notable:

   1. Back-to-Back Frame benchmark was very consistent for some fixed
       frame sizes, and somewhat variable for other frame sizes.

   2. The number of Back-to-Back Frames with zero loss reported for
       large frame sizes was unexpectedly long (translating to 30
       seconds of buffer time), and no explanation or measurement limit
       condition was indicated. It was important that the buffering
       time calculations were part of the referenced testing and
       analysis [VSPERF-b2b], because the calculated buffer time of 30
       seconds for some frame sizes was clearly wrong or highly suspect.
       On the other hand, a result expressed only as a large number of
       Back-to-Back Frames does not permit such an easy comparison with
       reality.

   3. Calculation of the extent of buffer time in the DUT helped to
       explain the results observed with all frame sizes. For example,
       tests with some frame sizes cannot exceed the frame-header-
       processing rate of the DUT, thus, no buffering occurs.
       Therefore, the results depended on the test equipment and not the
       DUT.

   4. It was found that a better estimate of the DUT buffer time could
       be calculated using measurements of both the longest burst in
       frames without loss and results from the Throughput tests
       conducted according to Section 26.1 of [RFC2544]. It is apparent
       that the DUT's frame-processing rate empties the buffer during a
       trial and tends to increase the "implied" buffer-size estimate
       (measured according to Section 26.4 of [RFC2544] because many
       frames have departed the buffer when the burst of frames ends).
       A calculation using the Throughput measurement can reveal a
       "corrected" buffer-size estimate.

   Further, if the Throughput tests of Section 26.1 of [RFC2544] are
   conducted as a prerequisite, the number of frame sizes required for
   Back-to-Back Frame benchmarking can be reduced to one or more of the
   small frame sizes, or the results for large frame sizes can be noted
   as invalid in the results if tested anyway. These are the larger
   frame sizes for which the Back-to-Back Frame rate cannot exceed the
   frame-header-processing rate of the DUT and little or no buffering
   occurs.

   The material below provides the details of the calculation to
   estimate the actual buffer storage available in the DUT, using
   results from the Throughput tests for each frame size and the Max

Theoretical Frame Rate for the DUT links (which constrain the minimum
frame spacing).

In reality, there are many buffers and packet-header-processing steps
in a typical DUT. The simplified model used in these calculations
for the DUT includes a packet-header-processing function with limited
rate of operation, as shown in Figure 1.

```
                    |------------ DUT --------|
Generator -> Ingress -> Buffer -> HeaderProc -> Egress -> Receiver
```
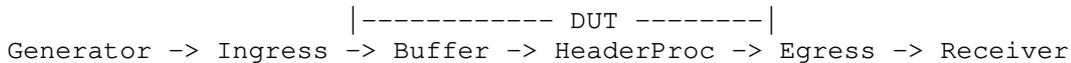
Figure 1: Simplified Model for DUT Testing

So, in the Back-to-Back Frame testing:

1. The ingress burst arrives at Max Theoretical Frame Rate, and
    initially the frames are buffered.

2. The packet-header-processing function (HeaderProc) operates at
    the "Measured Throughput" (Section 26.1 of [RFC2544]), removing
    frames from the buffer (this is the best approximation we have,
    another acceptable approximation is the received frame rate
    during Back-to-back Frame testing, if Measured Throughput is not
    available).

3. Frames that have been processed are clearly not in the buffer, so
    the Corrected DUT Buffer Time equation (Section 6.4) estimates
    and removes the frames that the DUT forwarded on egress during
    the burst. We define buffer time as the number of frames
    occupying the buffer divided by the Max Theoretical Frame Rate
    (on ingress) for the frame size under test.

4. A helpful concept is the buffer-filling rate, which is the
    difference between the Max Theoretical Frame Rate (ingress) and
    the Measured Throughput (HeaderProc on egress). If the actual
    buffer size in frames is known, the time to fill the buffer
    during a measurement can be calculated using the filling rate, as
    a check on measurements. However, the buffer in the model
    represents many buffers of different sizes in the DUT data path.

Knowledge of approximate buffer storage size (in time or bytes) may
be useful in estimating whether frame losses will occur if DUT
forwarding is temporarily suspended in a production deployment due to
an unexpected interruption of frame processing (an interruption of
duration greater than the estimated buffer would certainly cause lost
frames). In Section 6, the calculations for the correct buffer time
use the combination of offered load at Max Theoretical Frame Rate and
header-processing speed at 100% of Measured Throughput. Other
combinations are possible, such as changing the percent of Measured
Throughput to account for other processes reducing the header
processing rate.

The presentation of OPNFV VSPERF evaluation and development of
enhanced search algorithms [VSPERF-BSLV] was given and discussed at
IETF 102. The enhancements are intended to compensate for transient
processor interrupts that may cause loss at near-Throughput levels of
offered load. Subsequent analysis of the results indicates that
buffers within the DUT can compensate for some interrupts, and this
finding increases the importance of the Back-to-Back Frame
characterization described here.

5. Prerequisites

The test setup MUST be consistent with Figure 1 of [RFC2544], or
Figure 2 of that document when the tester's sender and receiver are
different devices. Other mandatory testing aspects described in
[RFC2544] MUST be included, unless explicitly modified in the next
section.

The ingress and egress link speeds and link-layer protocols MUST be
specified and used to compute the Max Theoretical Frame Rate when

respecting the minimum interframe gap.

The test results for the Throughput benchmark conducted according to
Section 26.1 of [RFC2544] for all frame sizes RECOMMENDED by
[RFC2544] MUST be available to reduce the tested-frame-size list or
to note invalid results for individual frame sizes (because the burst
length may be essentially infinite for large frame sizes).

Note that:

* the Throughput and the Back-to-Back Frame measurement-
   configuration traffic characteristics (unidirectional or
   bidirectional, and number of flows generated) MUST match.

* the Throughput measurement MUST be taken under zero-loss
   conditions, according to Section 26.1 of [RFC2544].

The Back-to-Back Benchmark described in Section 3.1 of [RFC1242] MUST
be measured directly by the tester, where buffer size is inferred
from Back-to-Back Frame bursts and associated packet-loss
measurements. Therefore, sources of frame loss that are unrelated to
consistent evaluation of buffer size SHOULD be identified and removed
or mitigated. Example sources include:

* On-path active components that are external to the DUT

* Operating-system environment interrupting DUT operation

* Shared-resource contention between the DUT and other off-path
   component(s) impacting DUT's behavior, sometimes called the "noisy
   neighbor" problem with virtualized network functions.

Mitigations applicable to some of the sources above are discussed in
Section 6.2, with the other measurement requirements described below
in Section 6.

6. Back-to-Back Frames

   Objective: To characterize the ability of a DUT to process Back-to-
   Back Frames as defined in [RFC1242].

   The procedure follows.

6.1. Preparing the List of Frame Sizes

   From the list of RECOMMENDED frame sizes (Section 9 of [RFC2544]),
   select the subset of frame sizes whose Measured Throughput (during
   prerequisite testing) was less than the Max Theoretical Frame Rate of
   the DUT/test setup. These are the only frame sizes where it is
   possible to produce a burst of frames that cause the DUT buffers to
   fill and eventually overflow, producing one or more discarded frames.

6.2. Test for a Single Frame Size

   Each trial in the test requires the tester to send a burst of frames
   (after idle time) with the minimum interframe gap and to count the
   corresponding frames forwarded by the DUT.

   The duration of the trial includes three REQUIRED components:

   1. The time to send the burst of frames (at the back-to-back rate),
       determined by the search algorithm.

   2. The time to receive the transferred burst of frames (at the
       [RFC2544] Throughput rate), possibly truncated by buffer
       overflow, and certainly including the latency of the DUT.

   3. At least 2 seconds not overlapping the time to receive the burst
       (Component 2, above), to ensure that DUT buffers have depleted.
       Longer times MUST be used when conditions warrant, such as when
       buffer times >2 seconds are measured or when burst sending times

are >2 seconds, but care is needed, since this time component
directly increases trial duration, and many trials and tests
comprise a complete benchmarking study.

The upper search limit for the time to send each burst MUST be
configurable to values as high as 30 seconds (buffer time results
reported at or near the configured upper limit are likely invalid,
and the test MUST be repeated with a higher search limit).

If all frames have been received, the tester increases the length of
the burst according to the search algorithm and performs another
trial.

If the received frame count is less than the number of frames in the
burst, then the limit of DUT processing and buffering may have been
exceeded, and the burst length for the next trial is determined by
the search algorithm (the burst length is typically reduced, but see
below).

Classic search algorithms have been adapted for use in benchmarking,
where the search requires discovery of a pair of outcomes, one with
no loss and another with loss, at load conditions within the
acceptable tolerance or accuracy. Conditions encountered when
benchmarking the infrastructure for network function virtualization
require algorithm enhancement. Fortunately, the adaptation of Binary
Search, and an enhanced Binary Search with Loss Verification, have
been specified in Clause 12.3 of [TST009]. These algorithms can
easily be used for Back-to-Back Frame benchmarking by replacing the
offered load level with burst length in frames. [TST009], Annex B
describes the theory behind the enhanced Binary Search with Loss
Verification algorithm.

There are also promising works in progress that may prove useful in
Back-to-Back Frame benchmarking. [BMWG-MLRSEARCH] and
[BMWG-PLRSEARCH] are two such examples.

Either the [TST009] Binary Search or Binary Search with Loss
Verification algorithms MUST be used, and input parameters to the
algorithm(s) MUST be reported.

The tester usually imposes a (configurable) minimum step size for
burst length, and the step size MUST be reported with the results (as
this influences the accuracy and variation of test results).

The original Section 26.4 of [RFC2544] definition is stated below:

| The back-to-back value is the number of frames in the longest
| burst that the DUT will handle without the loss of any frames.

6.3. Test Repetition and Benchmark

On this topic, Section 26.4 of [RFC2544] requires:

| The trial length MUST be at least 2 seconds and SHOULD be repeated
| at least 50 times with the average of the recorded values being
| reported.

Therefore, the Back-to-Back Frame benchmark is the average of burst-
length values over repeated tests to determine the longest burst of
frames that the DUT can successfully process and buffer without frame
loss. Each of the repeated tests completes an independent search
process.

In this update, the test MUST be repeated N times (the number of
repetitions is now a variable that must be reported) for each frame
size in the subset list, and each Back-to-Back Frame value MUST be
made available for further processing (below).

6.4. Benchmark Calculations

For each frame size, calculate the following summary statistics for

longest Back-to-Back Frame values over the N tests:

*   Average (Benchmark)

*   Minimum

*   Maximum

*   Standard Deviation

Further, calculate the Implied DUT Buffer Time and the Corrected DUT
Buffer Time in seconds, as follows:

Implied DUT buffer time =

   Average num of Back-to-back Frames / Max Theoretical Frame Rate

The formula above is simply expressing the burst of frames in units
of time.

The next step is to apply a correction factor that accounts for the
DUT's frame forwarding operation during the test (assuming the simple
model of the DUT composed of a buffer and a forwarding function,
described in Section 4).

```
Corrected DUT Buffer Time =
                    /                                           \
    Implied DUT    |Implied DUT       Measured Throughput        |
 =  Buffer Time -  |Buffer Time * --------------------------     |
                   |                Max Theoretical Frame Rate    |
                    \                                           /
```

   where:

1. The "Measured Throughput" is the [RFC2544] Throughput Benchmark
    for the frame size tested, as augmented by methods including the
    Binary Search with Loss Verification algorithm in [TST009] where
    applicable and MUST be expressed in frames per second in this
    equation.

2. The "Max Theoretical Frame Rate" is a calculated value for the
    interface speed and link-layer technology used, and it MUST be
    expressed in frames per second in this equation.

The term on the far right in the formula for Corrected DUT Buffer
Time accounts for all the frames in the burst that were transmitted
by the DUT *while the burst of frames was sent in*. So, these frames
are not in the buffer, and the buffer size is more accurately
estimated by excluding them. If Measured Throughput is not
available, an acceptable approximation is the received frame rate
(see Forwarding Rate in [RFC2889] measured during Back-to-back Frame
testing).

7. Reporting

    The Back-to-Back Frame results SHOULD be reported in the format of a
    table with a row for each of the tested frame sizes. There SHOULD be
    columns for the frame size and the resultant average frame count for
    each type of data stream tested.

    The number of tests averaged for the benchmark, N, MUST be reported.

    The minimum, maximum, and standard deviation across all complete
    tests SHOULD also be reported (they are referred to as
    "Min,Max,StdDev" in Table 1).

    The Corrected DUT Buffer Time SHOULD also be reported.

    If the tester operates using a limited maximum burst length in
    frames, then this maximum length SHOULD be reported.

```
+=============+================+================+================+
| Frame Size, | Ave B2B        | Min,Max,StdDev | Corrected Buff |
| octets      | Length, frames |                | Time, Sec      |
+=============+================+================+================+
| 64          | 26000          | 25500,27000,20 | 0.00004        |
+-------------+----------------+----------------+----------------+
```

                  Table 1: Back-to-Back Frame Results

   Static and configuration parameters (reported with Table 1):

   * Number of test repetitions, N

   * Minimum Step Size (during searches), in frames.

   If the tester has a specific (actual) frame rate of interest (less
   than the Throughput rate), it is useful to estimate the buffer time
   at that actual frame rate:

   Actual Buffer Time =
                                  Max Theoretical Frame Rate
       = Corrected DUT Buffer Time * --------------------------
                                       Actual Frame Rate

   and report this value, properly labeled.

8. Security Considerations

   Benchmarking activities as described in this memo are limited to
   technology characterization using controlled stimuli in a laboratory
   environment, with dedicated address space and the other constraints
   of [RFC2544].

   The benchmarking network topology will be an independent test setup
   and MUST NOT be connected to devices that may forward the test
   traffic into a production network or misroute traffic to the test
   management network. See [RFC6815].

   Further, benchmarking is performed on an "opaque-box" (a.k.a.
   "black-box") basis, relying solely on measurements observable
   external to the Device or System Under Test (SUT).

   The DUT developers are commonly independent from the personnel and
   institutions conducting benchmarking studies. DUT developers might
   have incentives to alter the performance of the DUT if the test
   conditions can be detected. Special capabilities SHOULD NOT exist in
   the DUT/SUT specifically for benchmarking purposes. Procedures
   described in this document are not designed to detect such activity.
   Additional testing outside of the scope of this document would be
   needed and has been used successfully in the past to discover such
   malpractices.

   Any implications for network security arising from the DUT/SUT SHOULD
   be identical in the lab and in production networks.

9. IANA Considerations

   This document has no IANA actions.

10. References

10.1. Normative References

   [RFC1242] Bradner, S., "Benchmarking Terminology for Network
              Interconnection Devices", RFC 1242, DOI 10.17487/RFC1242,
              July 1991, <https://www.rfc-editor.org/info/rfc1242>.

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119,
              DOI 10.17487/RFC2119, March 1997,

              <https://www.rfc-editor.org/info/rfc2119>.

   [RFC2544] Bradner, S. and J. McQuaid, "Benchmarking Methodology for
              Network Interconnect Devices", RFC 2544,
              DOI 10.17487/RFC2544, March 1999,
              <https://www.rfc-editor.org/info/rfc2544>.

   [RFC6985] Morton, A., "IMIX Genome: Specification of Variable Packet
              Sizes for Additional Testing", RFC 6985,
              DOI 10.17487/RFC6985, July 2013,
              <https://www.rfc-editor.org/info/rfc6985>.

   [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC
              2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174,
              May 2017, <https://www.rfc-editor.org/info/rfc8174>.

   [RFC8239] Avramov, L. and J. Rapp, "Data Center Benchmarking
              Methodology", RFC 8239, DOI 10.17487/RFC8239, August 2017,
              <https://www.rfc-editor.org/info/rfc8239>.

   [TST009]   ETSI, "Network Functions Virtualisation (NFV) Release 3;
              Testing; Specification of Networking Benchmarks and
              Measurement Methods for NFVI", Rapporteur: A. Morton, ETSI
              GS NFV-TST 009 v3.4.1, December 2020,
              <https://www.etsi.org/deliver/etsi_gs/NFV-
              TST/001_099/009/03.04.01_60/gs_NFV-TST009v030401p.pdf>.

10.2. Informative References

   [BMWG-MLRSEARCH]
              Konstantynowicz, M., Ed. and V. Polák, Ed., "Multiple Loss
              Ratio Search for Packet Throughput (MLRsearch)", Work in
              Progress, Internet-Draft, draft-ietf-bmwg-mlrsearch-00, 9
              February 2021, <https://tools.ietf.org/html/draft-ietf-
              bmwg-mlrsearch-00>.

   [BMWG-PLRSEARCH]
              Konstantynowicz, M., Ed. and V. Polák, Ed., "Probabilistic
              Loss Ratio Search for Packet Throughput (PLRsearch)", Work
              in Progress, Internet-Draft, draft-vpolak-bmwg-plrsearch-
              03, 6 March 2020, <https://tools.ietf.org/html/draft-
              vpolak-bmwg-plrsearch-03>.

   [OPNFV-2017]
              Cooper, T., Rao, S., and A. Morton, "Dataplane
              Performance, Capacity, and Benchmarking in OPNFV", 15 June
              2017,
              <https://wiki.anuket.io/download/attachments/4404001/
              VSPERF-Dataplane-Perf-Cap-Bench.pdf?version=1&modification
              Date=1621191833500&api=v2>.

   [RFC1944] Bradner, S. and J. McQuaid, "Benchmarking Methodology for
              Network Interconnect Devices", RFC 1944,
              DOI 10.17487/RFC1944, May 1996,
              <https://www.rfc-editor.org/info/rfc1944>.

   [RFC2889] Mandeville, R. and J. Perser, "Benchmarking Methodology
              for LAN Switching Devices", RFC 2889,
              DOI 10.17487/RFC2889, August 2000,
              <https://www.rfc-editor.org/info/rfc2889>.

   [RFC5180] Popoviciu, C., Hamza, A., Van de Velde, G., and D.
              Dugatkin, "IPv6 Benchmarking Methodology for Network
              Interconnect Devices", RFC 5180, DOI 10.17487/RFC5180, May
              2008, <https://www.rfc-editor.org/info/rfc5180>.

   [RFC6201] Asati, R., Pignataro, C., Calabria, F., and C. Olvera,
              "Device Reset Characterization", RFC 6201,
              DOI 10.17487/RFC6201, March 2011,
              <https://www.rfc-editor.org/info/rfc6201>.

[RFC6815]   Bradner, S., Dubray, K., McQuaid, J., and A. Morton,
            "Applicability Statement for RFC 2544: Use on Production
            Networks Considered Harmful", RFC 6815,
            DOI 10.17487/RFC6815, November 2012,
            <https://www.rfc-editor.org/info/rfc6815>.

[VSPERF-b2b]
            Morton, A., "Back2Back Testing Time Series (from CI)", May
            2021, <https://wiki.anuket.io/display/HOME/
            Traffic+Generator+Testing#TrafficGeneratorTesting-
            AppendixB:Back2BackTestingTimeSeries(fromCI)>.

[VSPERF-BSLV]
            Rao, S. and A. Morton, "Evolution of Repeatability in
            Benchmarking: Fraser Plugfest (Summary for IETF BMWG)",
            July 2018,
            <https://datatracker.ietf.org/meeting/102/materials/
            slides-102-bmwg-evolution-of-repeatability-in-
            benchmarking-fraser-plugfest-summary-for-ietf-bmwg-00>.

[VSPERF-CI]
            Tahhan, M., "OPNFV VSPERF CI", September 2019,
            <https://wiki.anuket.io/display/HOME/VSPERF+CI>.

Acknowledgments

Author's Address

   Al Morton
   AT&T Labs
   200 Laurel Avenue South
   Middletown, NJ 07748
   United States of America

   Phone: +1 732 420 1571
   Email: acmorton@att.com