White Paper on Pseudo-ESN Collisions

TIA TR-45 ESN/UIM Ad Hoc Group
Gary Pellegrino – CommFlow Resources, Chair

Frank Quick – QUALCOMM Incorporated, Editor

26 May 2005

1 COLLISIONS

Two types of collisions are considered in this paper:

- Collisions in the addressing of messages on the CDMA forward common channel.
- Collisions in the long codes used for various purposes in the CDMA system.

Both of these types of collisions can result when two mobiles have the same 32-bit ESN. Although ESNs have always been unique in the past, the introduction of the 56-bit MEID has let to creation of a 32-bit "pseudo-ESN" (pESN) for backward compatibility with legacy systems. Since the pESN cannot be uniquely assigned, these types of collisions can occur between MEID-equipped phones.

This paper discusses the effects of such collisions and the frequency at which they are likely to occur.

1.1 pESN Collision Probability and SHA-1

The pESN is formed from a fixed, 8-bit "manufacturer code" (hexadecimal 80) and a 24-bit pseudorandom number. The SHA-1 hash function is used to hash the MEID to produce this pseudo-random number. The desirable property of the algorithm in this usage is that for any two distinct MEIDs chosen at random, the probability of producing the same pESN is no greater than chance (i.e., one in 2^{24}.) SHA-1 was chosen because it was believed to be the best available algorithm at the time.

Researchers have recently described a method for forming two message texts that produce the same SHA-1 hash. This attack may have important consequences for some applications of SHA-1, but it has no effect on the property used in the generation of pESN. Thus there is no reason to replace SHA-1 for the generation of pESN.

It should be noted that the generation of pESN is not a "secure" process: the probability of collision for pESNs is already high enough to allow the identification of colliding MEIDs by a brute-force search.

1.2 Address Collisions

On the CDMA paging channels, messages can be addressed by ESN. If two mobiles have the same pESN, and both receive an ESN-addressed message sent to either, both may accept and process the message.

For example, an SMS message sent by ESN could be received by both mobiles.

On the other hand, channel assignment messages sent by ESN should not be processed by both mobiles unless both attempt system access at the same time. While possible, this should be an infrequent event.

Paging records do not include an ESN address type (only IMSI or TMSI), hence pages should be received only by the intended mobile.

To avoid these types of collisions, ESN addressing should not be used for most individually addressed messages on the paging channels. Some messages, such as channel assignments, could be addressed by ESN if the collision rate for system accesses is acceptably low.

The following table lists the mobile-directed messages in TIA/EIA/IS-2000-A-1 and the expected consequences if received and processed by two mobiles:

Message	Consequences
Abbreviated Alert Order	Both phones may alert the user.
Audit Order	Unknown
Authentication Challenge Message	Both mobiles will respond. Denial of service is a possible result.
Base Station Challenge Confirmation Order	See SSD Update Message.
Channel Assignment Message, Extended Channel Assignment Message	Usually no effect: Traffic channel assignment is ignored by a mobile not in Origination or Page Response substates. Traffic Channel collision will occur if both mobiles are in these substates. Paging channel assignment would transition both mobiles to a new paging channel. One mobile may be unable to receive incoming calls.
Data Burst Message	For SMS, results in the message being received by both phones and displayed to both users. For special SMS commands (e.g., BREW application commands, push service emulation) the effects are unpredictable.
Feature Notification Message	Both phones will display the feature status. E.g., both will show voicemail even though only one has voicemail. Likely result is customer complaint.
General Page Message, Universal Page Message	No effect (not addressable by ESN)
Intercept Order	Usually no effect: not processed in Idle state; if both mobiles are in System Access state, both give intercept indication, could result in call drop for both.
Local Control Order	Unknown
Lock Until Power-Cycled Order	Both phones will be locked. Likely result is customer complaint.
Maintenance Required Order	Unknown

PACA Message	Both mobiles will respond, one likely with a Reject Order. The effect in the network is unpredictable.
Registration Accepted Order	Both mobiles will process the message. Could result in incorrect roaming indicator and/or incorrect common channel encryption key in one mobile.
Registration Rejected Order	Both mobiles will process the message. Results in loss of TMSI in one mobile, may result in denial of service.
Registration Request Order	Both mobiles will attempt to register. By itself this has no serious effect.
Release Order	Usually no effect: not processed in Idle state; if both mobiles are in System Access state, could result in call drop for both.
Reorder Order	Usually no effect: not processed in Idle state; if both mobiles are in System Access state, both give reorder indication, could result in call drop for both.
Retry Order	Usually no effect: not processed in Idle state; if both mobiles are in System Access state, both give busy indication, could result in call drop for both.
Security Mode Command Message	Both mobiles will process the message. Could result in incorrect security mode for one mobile.
Service Redirection Message	Both mobiles will process the message. Denial of service may result.
Service Release Message	Unknown
Slotted Mode Order	Both mobiles will process the message. One mobile may operate on an incorrect slot cycle and may be unable to receive further directed messages.
SSD Update Message	Both phones will initiate SSD Update. SSD Update will fail for one or both phones when the Base Station Challenge Response is received. This condition could persist, resulting in an inability to update SSD. Denial of service could also result.
Status Request Message	Both phones will respond. Effect in the network is unpredictable.
TMSI Assignment Message	Both phones will receive the same TMSI, causing collisions on all further messaging addressed by TMSI.
Unlock Order	Usually no effect: ignored by a phone not locked.
User Zone Reject Message	Unknown

1.3 Long Code Collisions

Long code collisions may occur when two mobiles located in the same sector and same CDMA channel use the same long code mask.

1.3.1 Access Channel Long Code Collision Effects

Long code collisions occur regularly, by design, on the CDMA Access Channels. Here the protocol uses a prescribed retransmission procedure to overcome the effects of collisions. These procedures include PN randomization and a retransmission protocol using a pseudorandom backoff similar to that employed in Ethernet.

PN randomization applies to access probe timing. Each probe is delayed relative to the start of the reverse link frame by a number that can be from 0 to 511 chips. This helps resolve long code collisions on the access channel by delaying the scrambling code, making each arriving access probe appear as a distinct correlation peak allowing separate demodulation by individual demodulation fingers. The exact delay is a hash of the ESN over the maximum delay specified in the access parameters message.

Since the maximum delay is at most 9 bits in length, ordinary ESNs already collide with minimum likelihood 1/512, which is much greater than the probability of a collision between two pESNs. Further, collisions are persistent in the sense that two ESNs that hash to the same PN delay always do so.

The retransmission backoff protocol, however, reduces the probability of sequential collisions by spreading the access probes over multiple frames. This is sufficient to mitigate PN randomization collisions in current systems and should remain effective with pESN.

1.3.2 Traffic Channel Long Code Collision Effects

On the CDMA traffic channels, collisions were not intended to occur, and no special protocol has been created to allow for collisions. So long as the mobile ESN is unique, Public Long Code Mask collisions are prevented. However, collisions can occur when using the Private Long Code Mask, since that mask is generated pseudo-randomly and therefore may not be unique. The probability of a Private Long Code Mask collision has been assumed to be small enough to be acceptable. Further, the Private Long Code Mask is changed on each call attempt, so such rare collisions should only result in a single call failure.

With the introduction of MEID and the derived pseudo-ESN, Public Long Code Mask collisions are possible. Since the probability of collision with pseudo-ESN is much greater than that for Private Long Code Mask, there is a concern that such collisions may occur more frequently. Moreover, the Public Long Code Mask derived from pseudo-ESN is static, so collisions could result in multiple call failures.

Long code mask collisions do not affect the forward traffic channels, because the calls are assigned to distinct Walsh code channels. Reverse traffic channels, however, are adversely affected, since calls are distinguished only by the long code mask.

If two calls in the same sector (or neighboring sectors in some cases) use the same long code mask and radio configuration, the base station(s) will set up two radio receiving elements with identical configurations. In the same sector, these elements will perform exactly the same demodulation processes and will obtain identical results. The two mobiles will be sending distinct frame data, but the base station receiving elements can receive at most one good frame per 20 msec, and a good frame can result only when one mobile's signal dominates at the antenna.

If the two signals arrive at the antenna with about the same power level, it is most likely that the two signals will interfere destructively and both calls will fail. Until one or the other of the calls drops, the interference can cause a high frame error rate. This may cause the base stations to increase the power control set point for each of the two mobiles, resulting in an increase in the reverse link interference caused by those mobiles, decreasing system capacity. The amount of the interference increase is dependent on the network implementation. A robust implementation would not allow the power levels to increase without bound, so in practice the amount of capacity loss is limited and should be of relatively short duration.

More likely, however, the two signals will arrive with different power levels, and both base station receivers will receive the frames from the dominant signal. This means the mobile users will hear the correct forward audio from their distinct calls, but the land-sides will hear the same reverse audio from the dominant mobile. We believe this has been demonstrated by KDDI.

Thus, for the time period that both calls remain active, the consequences are to cross connect the reverse audio from one mobile to both land-side calls, and to block the reverse audio and signaling from the other mobile.

1.3.3 Duration of Cross-Connect

Two cases need consideration: where two colliding calls start at the same time, and when a colliding call begins during an established call.

1.3.3.1 Simultaneous colliding calls

In the first case, both calls may initially succeed if the assigned radio configurations and service options are identical. Although only one reverse link is being received, the “wrong” base station receiver may see what is expected and continue the call. If the signaling performed by the base station and mobiles is identical for the two calls, the two base station receivers may complete all negotiation and proceed to connect audio to the land-side call. This situation is more likely to occur on mobile-originated calls, since there is much less signaling before audio is connected.

If any distinct signaling message is sent to or from the non-dominant mobile, the response to that message will not be received, and the non-dominant call should drop. Also note that pilot strength measurements will not be received from the non-dominant call, and handoffs will be controlled by the dominant mobile only. Thus, if either phone moves out of the sector, the non-dominant call will drop.

We can conclude that initial cross-connect may occur if the radio configurations and service options are the same, and all initial signaling is identical. However, this scenario seems unlikely to occur in practice. It is more likely that this event will cause one of the calls to fail, and therefore will be experienced by the users as an increase in call failure rate.

1.3.3.2 Non-simultaneous colliding calls

For the second case, where an existing call is in progress when the colliding call begins, there will be no cross-connect unless the colliding call dominates, reasoning as follows:

If the existing call dominates, all initial signaling to/from the colliding mobile will fail, including the acknowledgment of the base station ack that is sent at the beginning of the call, and the colliding call will drop without affecting the existing call.

If the colliding call interferes with the existing call, both calls should drop.

If the colliding call dominates, then cross-connect is likely to occur until one or the other of the calls drops. Again, any signaling message to or from the non-dominant mobile will receive no response, and this should cause the non-dominant call to drop.

We can conclude that this scenario is the most likely source of problems: cross-connect may occur if the radio configurations and service options are the same and the colliding call is dominant over the existing call. Such collisions can occur and will persist until signaling fails to the non-dominant mobile, or the callers end the call.

From the non-dominant callers' point of view, suddenly another call is injected into theirs (somewhat similar events occurred frequently with AMPS), after which they will probably hang up. The dominant caller is unaware that the reverse audio is cross-connected. If the dominant caller continues the call, the non-dominant mobile is effectively blocked, so the call cannot be reconnected.

Note that it is also possible for cross-connect to occur during handoff. This, however, requires that a dominant mobile appear in a sector with a colliding non-dominant mobile. This would only happen when the intruding mobile makes a very rapid transition into the sector (e.g., moving rapidly around a terrain blockage) and becomes dominant before its interference causes the non-dominant call to drop. Such events may be important in certain locations but should not add appreciably to the probability of cross-connect in the system as a whole.

2 PROBABILITY OF COLLISION

2.1 Introduction

We are interested in calculating the daily expected collisions an operator's network would experience if handsets using pseudoESN for PLCM were used.. Although, we can only get a ballpark figure for number of collisions expected, we want to provide as much realism as possible by using actual occupancy and call traffic experienced by a typical sector rather than assuming that the maximum number of calls are active all the time in a sector. Below we present the resulting formula and the parameters used in its calculations; they are presented for hourly intervals and the daily collision can be calculated by summing over the 24 hourly intervals. The derivation of the formula and its justification are given later in the paper. EQN 1 is a more detailed and accurate formula based on the call arrival rates in each sector. EQN 2 is an approximation to EQN 1 when only average sector call arrival rate is known.

E[daily collisions in the network] = $\sum_{h=1}^{24}$ E[collisions in hour h]

E[hourly collisions in the network] $\approx \frac{f^2(I+1)}{N\mu} \sum_{i=1}^{U} \lambda_i^2$ (EQN 1)

$\approx \frac{2f^2(I+1)}{N\mu} U \lambda^2$ (EQN 2)

2.2 Definitions

U: the total number of carrier sectors[1] in the operator's universe.

λ_i: expected number of calls made per hour h in a sector i.

λ: expected number of calls made per hour h in an average sector.

1/μ: average call holding time.

f: fraction of the calls that use pseudoESN based handsets

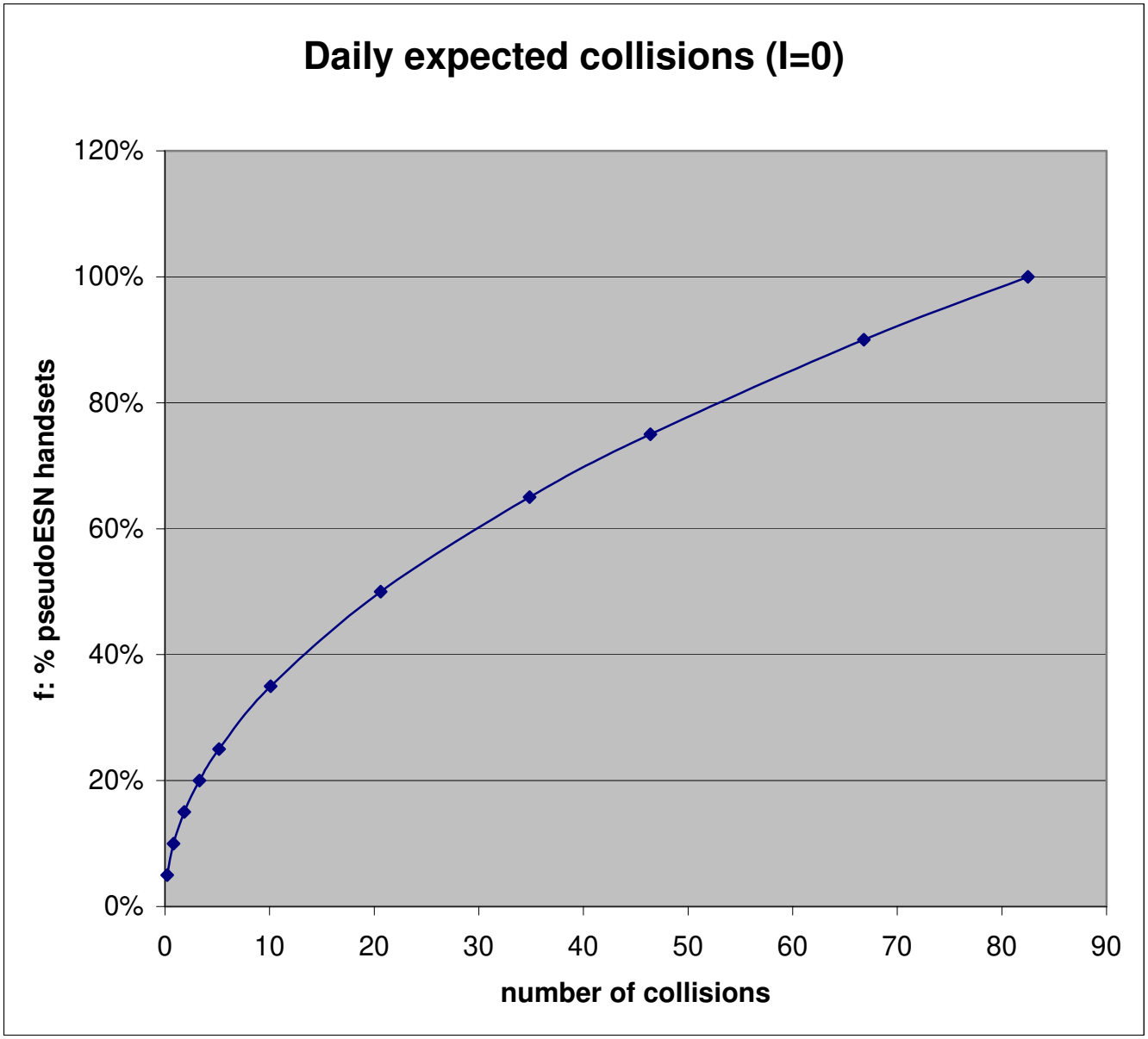
I: number of neighboring sectors that interfere with the current sector.

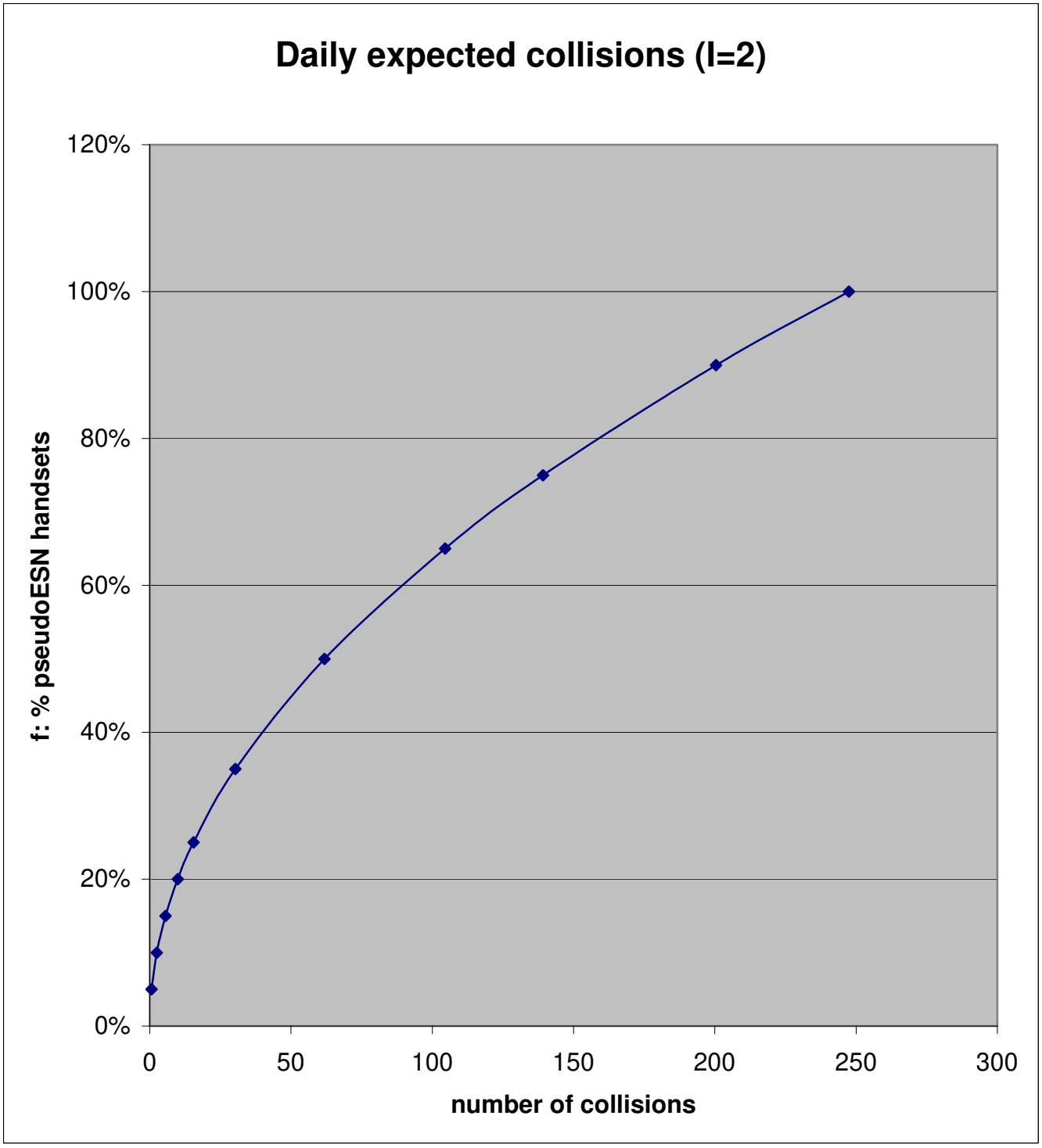
N: the possible values for the public long code mask (PLCM): 2^{24} with pseudoESNs.

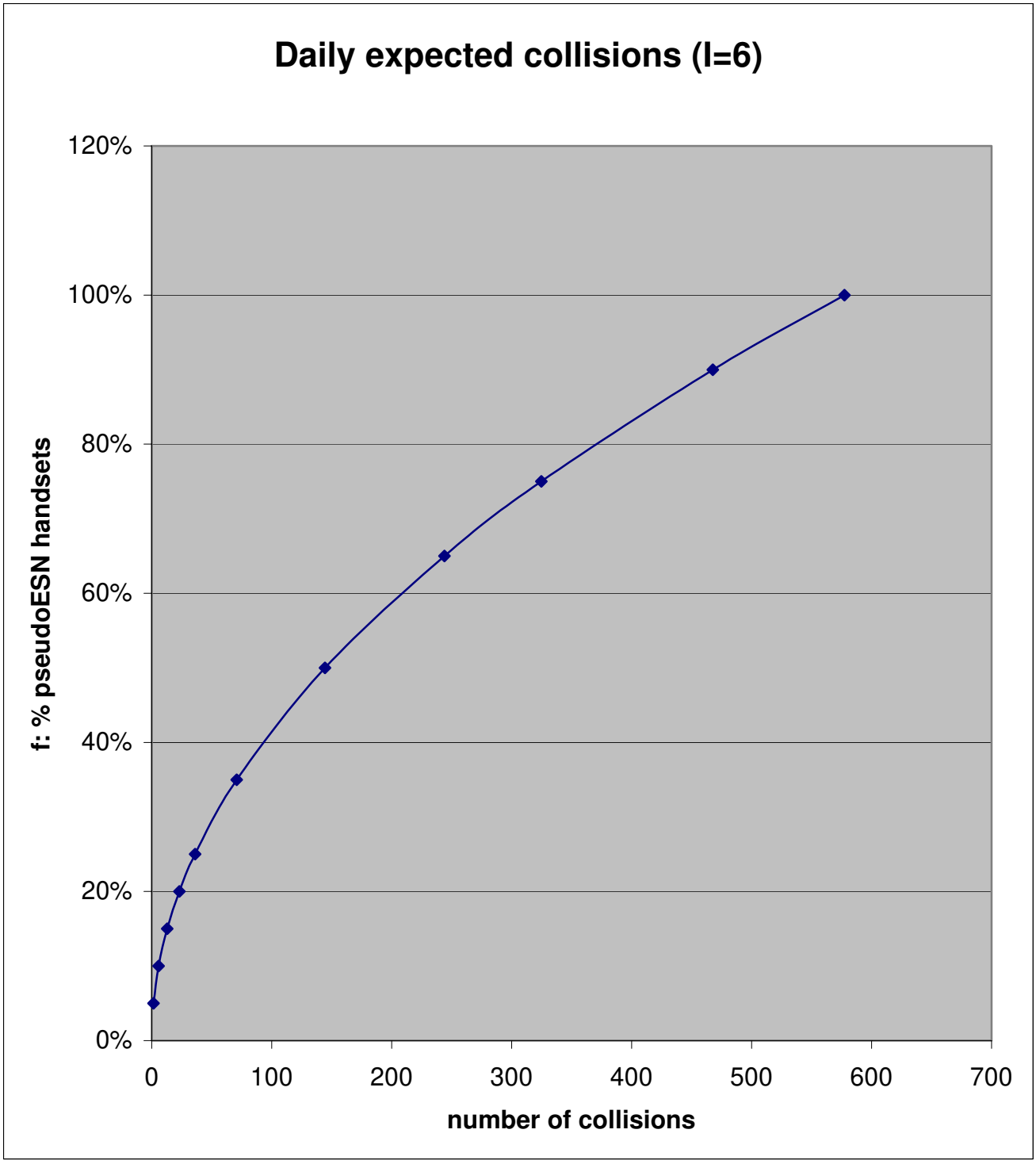
2.3 Example Results

For I=0,2,6 interfering neighboring sectors we plot the graph for the daily expected collisions (using EQN 2) as they vary when f ranges from 5 to 100%. U=60,000 and λ is calculated from example data described in a section below.

[1] When we use "sector" we mean carrier-sector in the rest of this section.


Daily expected collisions (I=0)
f: % pseudoESN handsets
120%
100%
80%
60%
40%
20%
0%
0
10
20
30
40
50
60
70
80
90
number of collisions


Daily expected collisions (I=2)
f: % pseudoESN handsets
120%
100%
80%
60%
40%
20%
0%
0
50
100
150
200
250
300
number of collisions



2.4 Derivation of the formula

E[daily collisions in network]

$= E\sum_{h=1}^{24}$ (collisions in hour h) $= \sum_{h=1}^{24}$ E[collisions in hour h]

/* by linearity of expectation*/

E[hourly collisions in network]

$= E\sum_{i=1}^{U}$ (hourly collisions due to calls in sector i)$= \sum_{i=1}^{U}$ E[hourly collisions due to calls in sector i]

$= \sum_{i=1}^{U}$ (E[collisions due to other calls in current sector i] + E[collisions due to other calls in neighboring sectors])

$= \sum_{i=1}^{U}$ (E[collisions due to other calls in current sector i] + $\sum_{j=1}^{I}$ E[collisions due to other calls in neighboring sector j])

$= \sum_{i=1}^{U}$ (E[collisions due to calls in current sector i] + I*E[collisions due to other calls in neighboring sector j]) /* by linearity of expectation */

E[hourly collisions due to other calls in neighboring sector]

= E[# of collisions with calls in neighboring sector when 1st call arrives in current sector +…+ # of collisions with calls in neighboring sector when last call arrives in current sector]

= E[# call carried in current sector i]*E[collisions with neighboring sector when a call arrives in current sector] /* due to independence*/

/*Let χj be the indicator variable which is 1 if collision happens between the new call in the current sector and the i-th call in the neighboring sector & 0 otherwise*/

/* m be the simultaneous calls in the neighboring sector*/

$= \lambda i*E[\chi_1 + \chi_2 +\ldots+ \chi_m] = \lambda i*E[m]*E[\chi] = \lambda i*E[m]*P[\chi=1]$

$= \lambda i*E[m]*(1/N) = (\lambda i/N)*((\lambda j/\mu)(1-P_B)) \leq (\lambda i/N)(\lambda j/\mu)$

/* since $E[m] = (\lambda/\mu)(1-P_B)$ due to Erlang B distribution*/

/*since blocking prob is very low, (< 1%), the approximation above is good */

/* assuming neighboring sector have similar traffic, then $\lambda i=\lambda j$ and it is an upper bound also */

$\approx 1/N\mu * \lambda_i^2$

E[hourly collisions due to other calls in the current sector i]

/* Let Cj be the # of calls that arrive given that there are already j calls in the system*/

= E[C1*(# of collisions when 1 call in system) + C2*(# of collisions when 2 calls in system) +…+Cm-1*(# of collisions when m-1 calls in system)]

/* here m is the maximum number of calls possible in the system */

$= E[\sum_{j=1}^{m-1} Cj$ *(# of collisions with j calls in system and a new call arrives)]

$= E[\sum_{j=1}^{m-1} Cj$ *E[χ1 + χ2 +…+ χj|j calls]] = $E[\sum_{j=1}^{m-1} Cj$ *j*E[χ]]= $E[\sum_{j=1}^{m-1} Cj$ *j*P[χ=1]] =

$E[\sum_{j=1}^{m-1} Cj$ *j*(1/N)] = $\sum_{j=1}^{m-1} E[Cj]$ *j*(1/N) = $\sum_{j=1}^{m-1} Pj$ *λ_i *j*(1/N)

/* where χi is an indicator variable which is 1 if there is a collision between the new call and the existing i-th call */

/* E[Cj] = λ_i *Pj */

$=\lambda_i /N* \sum_{j=1}^{m-1} j * Pj = \lambda_i /N*(E[m] - m* P_B). \leq \lambda_i /N*E[m]$

/* where Pi is the probability that the system has i calls active*/

/* since E[m] for sector i = $(\lambda_i /\mu)(1-P_B)$ due to Erlang B distribution */

$\approx \lambda_i /N*(\lambda_i/\mu)(1-P_B) \leq 1/N\mu * \lambda_i^2$

hence the total E[hourly collisions in a network due to sector i]

$\approx 1/N\mu (\lambda_i^2+I*\lambda_i^2)$

hence the total E[hourly collisions in the network]

$$\approx \frac{(I+1)}{N\mu} \sum_{i=1}^{U} \lambda_i^2$$

/*since only f fraction of handsets are pseudoESN based, λ_i should be replaced by f*λ_i */

E[hourly collisions in the network]

$$\approx \frac{f^2(I+1)}{N\mu} \sum_{i=1}^{U} \lambda_i^2 \qquad \text{(EQN 1)}$$

The above equation EQN 1 requires the knowledge of the arrival rates at each sector. Perhaps we only know the average arrival rate taken over all sectors, in that case one can simply try to write the hourly collisions as $\frac{f^2(I+1)}{N\mu} U \lambda^2$, however, this may be a lower bound because may be the λ_i are distributed non-uniformly over the various sectors and the squaring makes this difference important. In that case the worst case for collisions would be if part of the sectors had maximum arrival rate and the rest of the sectors had zero arrival rate. This, however, would be an upper bound. We can try to find a middle ground by looking at the actual distribution of arrival rates across various sectors. Right now we will just try to get a reasonable bound by assuming that the average load seen by a sector is low and that the median sector is much below average. Then making sure things are consistent we get that in terms of the average arrival rate:

$$\text{E[hourly collisions in the network]} \approx \frac{2f^2(I+1)}{N\mu} U \lambda^2 \qquad \text{(EQN 2)}$$

The Equation 2 above would be different if the distribution of the calls in the sectors was very non-uniform; the formula for EQN 2 should be taken only as an example for the data in the next section. If real data is available then either EQN 1 should be used or a better approximation for EQN 2 should be used.

2.5 Sector traffic data

Below is a some example traffic data for a network listing the average call arrival rates for all 24 hourly segments. U, the number of sector-carriers, is set to 60,000; N is fixed at 2^{24}; average call holding time is fixed to 120 seconds. Then using our formula we are able to calculate the hourly collisions and add them to get the daily expected collisions. If more detailed data is available then perhaps better estimates and Equation 1 can be used.

Hour	Avg sect-carr hourly arrival rate
10	120
11	135
12	145
13	145
14	160
15	185
16	190
17	200

18	170
19	145
20	120
21	110
22	85
23	60
0	60
1	60
2	60
3	60
4	60
5	60
6	60
7	60
8	85
9	110

2.6 Discussion

The formula and the example charts show the dependency of the collisions on the fraction of the handsets that are using pseudoESN based PLCM. We see the square dependencies as we had expected. The formula implicitly takes into account the handoff rate assuming that the arrival rates we have in the data includes not only new origination and call termination attempts but also handoff when primary sector has changed. If that is not correct then the formula underestimates and should be adjusted to handle handoffs. For soft or softer handoff, we assume this is the same as interference from neighbors, and can be accounted for by appropriate use of I, the number of interfering neighbors; however, we hope that primary sector change indicates handset movements which needs to be accounted for and we hope is accounted for in the hourly call arrival rates. The handoff impact needs to be further considered.

2.7 Summary

With pseudo-ESN, it is likely that some collisions with cross-connect of reverse audio will occur unless some mitigation is implemented. The rate at which these collisions occur will depend on the rate of usage of MEID-equipped phones in legacy systems. Mitigation strategies should be investigated by the appropriate standards bodies.