Benefit of Network Coding for Probabilistic Packet Marking and Collecting Coupons from Different Perspectives at the Collector*

Dung Tien NGO(a), Tuan Anh LE(b), Nonmembers, Choong Seon HONG(c), Member, Sungwon LEE(d), Won-Tae LEE(e), and Jae-Jo LEE(f), Nonmembers

SUMMARY Probabilistic Packet Marking (PPM) is a scheme for IP traceback where each packet is marked randomly with an IP address of one router on the attack path in order for the victim to trace the source of attacks. In previous work, a network coding approach to PPM (PPM+NC) where each packet is marked with a random linear combination of router IP addresses was introduced to reduce number of packets required to infer the attack path. However, the previous work lacks a formal proof for benefit of network coding to PPM and its proposed scheme is restricted. In this paper, we propose a novel method to prove a strong theorem for benefit of network coding to PPM in the general case, which compares different perspectives (interests of collecting) at the collector in PPM+NC scheme. Then we propose Core PPM+NC schemes based on our core network coding approach to PPM. From experiments, we show that our Core PPM+NC schemes actually require less number of packets than previous schemes to infer the attack path. In addition, based on the relationship between Coupon Collector’s Problem (CCP) and PPM, we prove that there exists numerous designs that CCP still benefits from network coding.

key words: IP traceback, probabilistic packet marking, network coding, coupon collector’s problem, different perspectives

1. Introduction

The purpose of IP traceback [1] is to identify the actual source of attack packets. Probabilistic Packet Marking (PPM) [2] is a scheme for IP traceback, where routers on the attack path from attacker to victim randomly mark traversing packets with their IP addresses in order for the victim (collector) to infer the path of attack. In this scheme in Fig. 1, once a router decides to mark, it will overwrite the previous mark of the previous router on a packet. Therefore, each received packet at the collector contains at most one router’s mark. Thus, the collector’s problem in the PPM scheme is to collect all distinct markings of routers on the attack path, which is essentially a Coupon Collector’s Problem (CCP) [3, pp.32].

To reduce average number of packets required to infer the attack path, a network coding approach to PPM (PPM+NC) [4] was introduced, where each traversing packet on the attack path is marked randomly with a linear combination of some router IP addresses since the marking packet on the attack path is marked randomly with an IP address of one router. Thus, the collector’s problem in the PPM scheme is to collect all distinct markings of routers on the attack path, which is essentially a Coupon Collector’s Problem (CCP) [3, pp.32].

The limitation of [4] is that it does not give us a strong proof for the benefit of network coding to probabilistic packet marking. Specifically, [4] considered a simple PPM+NC scheme where all routers on the attack path have equal marking probabilities $p$ to mark on packets. In order to compare about quantity (i.e., average number of packets required to infer all routers on the attack path) between PPM+NC and PPM schemes, [4] based on an unpractical assumption in such restricted PPM+NC scheme: where once a router decides to mark a packet, it will result in, with high probability, an innovative linear combination (compared with ones contained in collected packets before). Besides, there is no formal proof for benefit of Practical PPM+NC scheme which is proposed from an analysis in an ideal PPM+NC scheme where IPv4 packet header is assumed to have enough space to store 32-bit router IP address. Strictly speaking, the benefit of such scheme is just showed via simulations.

In this paper, we use a novel method to prove a strong theorem for benefit of network coding to probabilistic packet marking in the general case where routers on the attack path could have any (equal or unequal) marking probabilities, and the IPv4 packet header could have any size to store a content of marking (in Sect. 3 and Sect. 4.1). To compare PPM and PPM+NC schemes, we compare two different perspectives (interests of collecting) U and D at the same collector in the PPM+NC where the collecting problem U of...
the collector in such PPM+NC scheme is equivalent to the collecting problem in PPM scheme, while D is the default collecting problem in PPM+NC scheme that the collector is interested in collecting packets until inferring all routers on the attack path. In addition to pointing out the limitation of Practical PPM+NC [4] in Appendix, we propose a core network coding approach to probabilistic packet marking and Core PPM+NC schemes in Sect.4. Via experiments and results, we show that our Core PPM+NC schemes actually require less average number of packets to infer all routers on the attack path than previous schemes.

Futhermore, network coding is just known to have benefit for CCP in the case every possible linear combinations of coupons has the same probability of receiving [5, pp.10–13]. In the next contribution in Sect. 5, based on our strong theorem for benefit of network coding to PPM, we prove that there exists numerous designs of network coding that still offer benefit to CCP.

2. Related Works

Among many proposed schemes [1], [6] for IP traceback, PPM scheme [2] is practical due to its low router overhead and incremental deployment support. There are some well known PPM-based schemes: Fragment Marking Scheme (FMS) [2], Advanced Marking Scheme (AMS) [7], and Fast Internet Traceback (FIT) [8]. They mainly use 16-bit IP identification field [2] in IPv4 packet header for marking, and could use more 8-bit TOS field and 1-bit fragment flag [9].

FMS [2] can do well in the single-path attack, but it has large number of false positives and high computation overhead in the multi-path attacks due to very large possible number of combinations of fragments marked at the same distance [7].

AMS [7] tackled FMS’s problems by assuming a map of upstream routers already built by traceroute tool before and using a set of hash functions instead of fragmentation to avoid gathering fragments, which reduce false positives and computation in the path reconstruction phase during attacks.

FIT [8] proposed using packet marking instead of traceroute tool in AMS to reduce false positives in the map of upstream routers. It also proposed a 1-bit distance mechanism (instead of well known 5-bit using) together with TTL modification technique, which enalbes large allocated space for marking leading to reduced false positives in the path reconstruction phase. However, FIT scheme always has false positives in the map reconstruction phase because FIT routers put its hash fragments in traversing packets, which impacts on the false positives in the path reconstruction phase.

In [4], Sattari et al. proposed a Practical PPM+NC scheme that combines random linear network coding [10] with PPM, where each marked packet received at the victim contains $k$ $b$-bit coefficients drawn uniformly at random from the Galois field $\mathbb{F}_q (q = 2^b)$ and an associated linear combination result of $k$ fragments with same offset from $k$ consecutive traceback routers. Simulations demonstrated that this scheme requires less average number of packets than FMS scheme to derive all routers on the attack path. However, its limitations are shown by our analysis in Appendix.

In this paper, we are interested in the metric reducing average number of packets required for the victim (collector) to infer all routers on the attack path. By our strong proof for benefit of network coding to PPM, we propose Core PPM+NC schemes which are better than previous schemes in terms of such metric.

3. Benefit of Network Coding for Probabilistic Packet Marking

In this section, we consider the case that IPv4 packet header has enough space to store an entire 32-bit IP address. For practical situations with limited size of IPv4 packet header leading to the fact that every router must divide its IP address into fragments to encode to traversing packets, we will consider in Sect. 4.1.

3.1 Background of Network Coding Approach to PPM Scheme

In network coding approach to PPM (PPM+NC) [4] shown in Fig. 2, the marking field in each packet’s header is divided into three fields: a random coefficient field of $k$ coefficient slots $c_1, \ldots, c_k$, a linear combination field to store a linear combination result $\sum_{i=1}^{k} c_i ID_i$ of $k$ router IP addresses $ID_1, \ldots, ID_k$ on the path of packet, and a distance field $dist$ to calculate the distance of the marking router from the collector. Once a router $i$ decides to mark a packet, it overwrites the previous mark on the packet: it zeros the entire marking field, and chooses a coefficient $c_i$ uniformly at random out of a field $\mathbb{F}_q = \{0, \ldots, q-1\}$, multiplies its IP address $ID_i$ with $c_i$ then write $c_i$ to the first slot of coefficient field and the result of $c_i ID_i$ to the linear combination field, finally resets the distance field $dist$ to zero. If the next router $j$ decides not to overwrite, it picks uniformly at random a coefficient $c_j$ out of $\mathbb{F}_q$ and write $c_j$ to the next coefficient slot, and adds the result of $c_j ID_j$ to the current content of the marking field by updating it to $c_i ID_i + c_j ID_j$, finally increments the distance field $dist$.

In the PPM+NC scheme introduce by [4] above, all routers on the attack path have equal marking probabilities $p$ to mark on packets. In addition, once marking, a router simply picks a coefficient uniformly at random from $\mathbb{F}_q$. Therefore, [4] simply considered benefit of network coding
in a restricted PPM+NC scheme. Another limitation of [4] is that authors used an unpractical assumption to simplify the difficulty in proving benefit of network coding for probabilistic packet marking: once a router decides to mark a packet in PPM+NC scheme, it will result in, with high probability, an innovative an innovative linear combination (compared with ones contained in collected packets before).

3.2 Strong Theorem for Benefit of Network Coding to Probabilistic Packet Marking

In this section, we compare PPM+NC with PPM scheme on a given path with any setting of marking probabilities \((P_m(n), \ldots, P_m(2), P_m(1))\) for \(n\) routers (Fig. 1 and Fig. 2). We consider general setting in PPM+NC scheme compared to [4]: Once a router decides to mark packet, it draws a coefficient randomly with any distribution over \(\mathbb{F}_q\) rather than drawing uniformly at random over \(\mathbb{F}_q\). In addition, our consideration for PPM+NC scheme is more general than Sattari et al.’s [4] because if the next router decides not to mark, it could pick a coefficient randomly with any distribution rather than uniformly at random from \(\mathbb{F}_q\).

3.2.1 Main Idea of Proof: Different Perspectives at the Collector

It is noticed that we should distinguish clearly between mechanisms (schemes) of filling routers’ information on traversing packets and interests (perspectives) of the collector: a collector could have many interests in the same mechanism, and each interest should be associated with a mechanism. In the same mechanism, different interests will require different number of packets to satisfy. Also, with the same interest, different mechanisms will have different number of packets to satisfy the collector.

Let \(X_{PPM,NC} \neq X_{PPM}\) be number of packets required to satisfy the default collector’s interests \(D\) inferring the attack path in PPM+NC and PPM schemes, respectively. Specifically, \(X_{PPM}\) is number of packets to receive enough \(n\) distinct types of marking in PPM scheme, whereas \(X_{PPM,NC}\) is number of packets to collect enough \(n\) independent linear combinations in PPM+NC scheme. It is difficult to compare such two independent and different stochastic schemes (PPM and PPM+NC schemes) with the same interests \(D\) of the collector. Our important idea is to find another interest (perspective) \(U\) in the PPM+NC scheme which has connection with PPM scheme then our problem of comparison between PPM+NC and PPM schemes is easier.

**Theorem 1:** There always exists a random variable \(X_U\) in PPM+NC scheme so that: \(X_U \neq X_{PPM}\) are i.i.d., and \(X_{PPM,NC} \leq X_U\).

Let \(X_U\) be number of packets required for the collector to collect \(n\) distinct markings of \(n\) routers on a given path in the PPM+NC scheme. Note that the common point between PPM+NC and PPM schemes is that each marking decision by any router on the attack path does not depend on the running scheme (PPM+NC or PPM) of \(n\) routers, while the only difference between them are the content of marking on each packet. Therefore, the number of packets required for the collector to collect \(n\) distinct markings of \(n\) routers on the given path does not depend on the running scheme of \(n\) routers. As a result, \(X_U\) and \(X_{PPM}\) are two independent and identically distributed (i.i.d.) random variables. This reveals the connection between PPM+NC and PPM schemes.

From now, in order to compare \(E[X_{PPM,NC}]\) in the PPM+NC scheme and \(E[X_{PPM}]\) in the PPM scheme, we just need to compare \(X_{PPM,NC}\) and \(X_U\) which correspond to two different perspectives (interests of collecting) \(D\) and \(U\) of the collector in the PPM+NC scheme. Each time receiving a new packet in the PPM+NC scheme, the collector checks whether his interest \(U\) or \(D\) is satisfied first. Based on the value of distance field \(\text{dist}\) of the received packet, the collector determines its distance \(i\) from the marking router then derive that \(k\) coefficients \((c_1, c_2, \ldots, c_{i+k-1})\) stored in the random coefficient field are drawn randomly and independently by routers at distances \(i, i-1, \ldots, i-k+1\). Thus, each received packet corresponds to a coefficient vector \((c_1, c_2, \ldots, c_{i+k-1})\) and its corresponding linear combination result \(\sum_{j=i}^{i+k-1} c_j ID_j\). Because packets marked at different distances correspond to coefficient vectors that belong to different space of vectors while the goal of collector is to derive \(n\) router IDs on the attack path, it is reasonable to unify the representation for every received packet. As a result, every received coefficient vector \((c_1, c_2, \ldots, c_{i+k-1})\) could be represented by the corresponding coefficient vector \((c_n, c_{n-1}, \ldots, c_1)\) in the \(n\)-dimensional space \(\mathbb{F}_q^n\) such that the coefficients \(c_j\) corresponding to routers at the remaining distances \(j\) with \((j > i) \lor (j < i-k+1)\) on the attack path are zeros. With this representation, \(X_{PPM,NC}\) is interpreted as the number of packets required for the collector to collect \(n\) linearly independent vectors from the \(n\)-dimensional space \(\mathbb{F}_q^n\).

To compare \(X_U\) with \(X_{PPM,NC}\), our idea is to establish a setting in PPM+NC scheme so that it transforms \(X_U\) into number of packets required to collect some \(n\) independent linear combinations from \(\mathbb{F}_q^n\), then we can easily compare \(X_U\) with \(X_{PPM,NC}\) in terms of collecting \(n\) linearly independent vectors. Let us define \(A_i\) as follows:

\[
A_i = \{c \in \mathbb{F}_q^n : (c[i] \neq 0) \land (c[j] = 0, (j > i)), \quad (1 \leq i \leq n)\}
\]

In [4], each packet marked by the \(i\)th router on the attack path corresponds to a vector that belongs to \(A_1 \cup A_2 \cup \ldots \cup A_i\) because the marking router and next routers on the attack path could pick zero coefficients from \(\mathbb{F}_q\) to put onto that packet. With such setting, \(n\) collected distinct markings could be not correspond to \(n\) linearly independent vectors, which make us difficult to compare \(X_U\) and \(X_{PPM,NC}\) in terms of collecting \(n\) linearly independent vectors. For example, the first packet is marked by the 2nd router on the attack path but that router picks a zero coefficient, while the second packet is marked by the 1st router (the nearest router of the victim) with non-zero coefficient from \(\mathbb{F}_q\). Clearly, they are not linearly independent vectors because the ma-
matrix \[
\begin{bmatrix}
0 & \cdots & 0 & * \\
0 & \cdots & 0 & * \\
0 & \cdots & 0 & *
\end{bmatrix}
\] formed by such two vectors is not an echelon matrix. However, we notice a fact from linear algebra [11, pp.170] that any \(n\) vectors which form an upper triangular matrix are linearly independent, i.e., such \(n\) vectors belong to \(n\) sets \(A_n, A_{n-1}, \ldots, A_1\) respectively. This leads us to establish the setting in PPM+NC scheme to ensure that \(n\) collected distinct markings corresponds to \(n\) linearly independent vectors: once marking, the marking router always picks a non-zero coefficient from \(\mathbb{F}_q\}\{0\}. This is our important setting and trick to compare \(X_U\) and \(X_{PPM+NC}\) easily in terms of collecting \(n\) linearly independent vectors in PPM+NC scheme. In addition, with this setting, we could prove the benefit of network coding for the general case of PPM+NC scheme where the marking router and next routers (on the attack path of a packet) pick a coefficient randomly with any distribution from \(\mathbb{F}_q\}\{0\}\) and \(\mathbb{F}_q\), respectively.

In our considered PPM+NC scheme where the marking router always picks a non-zero coefficient from \(\mathbb{F}_q\}\{0\}, a received packet marked by a router at distance \(i\) from the collector on the attack path corresponds to a coefficient vector that belongs to the set \(A_i \subset \mathbb{F}_q^n\). Because \(A_n = \{c \in \mathbb{F}_q^n : c[n] \neq 0\}\),

\[\Rightarrow \mathbb{F}_q^n - A_n = \{c \in \mathbb{F}_q^n : c[n] = 0\} .\]

Besides, \(A_{n-1} = \{c \in \mathbb{F}_q^n : (c[n] = 0) \wedge (c[n-1] \neq 0)\}\). Thus:

\[
\begin{cases}
A_{n-1} \subset \mathbb{F}_q^n - A_n \\
A_{n-1}, A_n \text{ are mutually disjoint.}
\end{cases}
\]

Doing the same steps above, we derive:

\[
\begin{aligned}
A_i & \subset (\mathbb{F}_q^n - A_n - A_{n-1} - \ldots - A_{i+1}) \\ A_i, A_{i+1}, \ldots, A_n & \text{ are mutually disjoint (} 1 \leq i \leq n-1) \\ \{0^n\} & = \mathbb{F}_q^n - A_n - A_{n-1} - \ldots - A_1.
\end{aligned}
\]

Note that a received packet not marked by any router on the attack path corresponds to the zero-coefficient vector \(\{0^n\}\). Hence, each received packet at the collector corresponds to a coefficient vector that belongs to one of \(n+1\) mutually disjoint sets \(A_n, A_{n-1}, \ldots, A_1, \{0^n\}\), where \(\mathbb{F}_q^n = A_n \cup A_{n-1} \cup \ldots \cup A_1 \cup \{0^n\}\). As a result, \(X_U\) is interpreted as the number of packets required to collect \(n\) coefficient vectors \(a_n, a_{n-1}, \ldots, a_1\) from \(n\) mutually disjoint sets \(A_n, A_{n-1}, \ldots, A_1\), respectively. The important point is that once the collector collected such \(n\) coefficient vectors, they would form an upper triangular matrix \(U\):

\[
U = \begin{bmatrix}
da_n \\
da_{n-1} \\
\vdots \\
da_1
\end{bmatrix} = \begin{bmatrix}
\bullet & * & * & * \\
0 & * & * & * \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \bullet
\end{bmatrix}.
\]

Clearly, \(\text{rank}(U) = n\). Thus, \(n\) coefficient vectors \(a_n, a_{n-1}, \ldots, a_1\) (row vector form) are linearly independent vectors. In other words, \(X_U\) is the number of packets required for the collector to collect \(n\) linearly independent vectors from the \(n\)-dimensional space \(\mathbb{F}_q^n\) so that such \(n\) vectors belong to \(n\) mutually disjoint sets \(A_n, A_{n-1}, \ldots, A_1\), respectively. Besides, we recall that \(X_{PPM+NC}\) is the minimum number of packets required for the collector to collect \(n\) linearly independent vectors from the \(n\)-dimensional space \(\mathbb{F}_q^n\). Therefore, for incoming packets at the collector, it is obviously that the first interest of collecting (which corresponds to \(X_{PPM+NC}\)) is satisfied before or at the same time the second interest of collecting (which corresponds to \(X_U\)) is satisfied:

\[X_{PPM+NC} \leq X_U .\]

Theorem 1 gives us an insight: The difference between collecting in PPM+NC scheme and PPM scheme could be interpreted by two collecting problems of the collector in PPM+NC scheme, that is, collecting any \(n\)-rank matrix in the PPM+NC scheme compared with collecting a \(n\)-rank upper triangle matrix. That is the critical reason leading to the benefit of network coding.

**Corollary 2:** \(E[X_{PPM+NC}] \leq E[X_{PPM}]\)

Corollary 2 proves the benefit of network coding to probabilistic packet marking by proving that PPM+NC scheme requires less expected number of packets than PPM scheme to derive all \(n\) router IDs on the attack path.

From Theorem 1, \(X_U\) and \(X_{PPM}\) are i.i.d. random variables. Therefore, \(E[X_U] = E[X_{PPM}]\).

On the other hand:

\[X_{PPM+NC} \leq X_U \Rightarrow E[X_{PPM+NC}] \leq E[X_U] .\]

Therefore:

\[E[X_{PPM+NC}] \leq E[X_{PPM}] .\]

3.3 Application of Different Perspectives Technique in Simulations

Let \(\overline{X}_{PPM+NC}\) (\(\overline{X}_{PPM}\) respectively) be the average number of packets required for the collector to derive all \(n\) router IDs on the attack path in the PPM+NC (PPM respectively) scheme among \(N\) independent experiments. Without theorem 1, corollary 2 according to the strong law of large numbers [12, Theorem 2.1] just says that once the number of independent experiments \(N\) in each scheme is sufficiently large, \(\overline{X}_{PPM+NC} \leq \overline{X}_{PPM}\). It means that, the simulation for checking such result must be run with \(2N\) experiments in total: \(N\) experiments in PPM+NC scheme, \(N\) experiments in PPM scheme. While, the interesting point from the Theorem 1 says that instead of running both two scheme PPM+NC and PPM scheme to compare \(\overline{X}_{PPM+NC}\) and \(\overline{X}_{PPM}\), we just need to run only \(N\) experiments in only one PPM+NC scheme then compare different perspectives (interests of collecting) \(\overline{X}_{PPM+NC}\) and \(\overline{X}_U\) of the collector.
where \( X_U \) is the average number of packets required for the collector to collect \( n \) distinct markings of \( n \) routers on the attack path in the PPM+NC scheme among \( N \) independent experiments.

### 4. Core PPM+NC Schemes

In this section, after proving the benefit of network coding to PPM in practical situations with limited budget in IPv4 packet header in Sect. 4.1, we propose a core network coding approach to PPM in Sect. 4.2. Then, we propose Core PPM+NC schemes in Sects. 4.3 and 4.4. Through experiments, we show that our Core PPM+NC schemes actually require less average number of packets for the victim to infer the attack path.

#### 4.1 Benefit of Network Coding to PPM in General Case

In practice, IPv4 packet header has not enough space to contain 32-bit router IP address which leads to the fact that every router must divide its IP address into fragments to encode to traversing packets. That is, number of fragments per router is \( f > 1 \). In [4], there is no formal proof for benefit of network coding for PPM+NC in such general case. Therefore, we will give our proof in this subsection.

**Theorem 3:** For any design of PPM+NC scheme taking linear combination of any \( k \) fragments (including at least one fragment of the marking router) from routers on the path from the marking router to the victim, that scheme always requires less expected number of packets than PPM scheme.

Is there any connection between cases \( f > 1 \) and \( f = 1 \)? Can we reuse results from Theorem 1 and Corollary 2? Our following analysis can tackle that problem. Based on the fact that if we collect \( nf \) fragments together with information about their offsets and distances then we, of course, infer all \( n \) router IP addresses. Therefore, PPM+NC scheme in the case \( f > 1 \) could be formulated as follows:

Each marked packet corresponds to one of \( nf \) distinct types of markings. Setting of \( nf \) marking probabilities are:

\[
P_m(n, 1), \ldots, P_m(n, f), \ldots, P_m(1, 1), \ldots, P_m(1, f)
\]

\( f \) fragments of the \( n \)th router \( f \) fragments of the 1st router

Because Theorem 1 holds for any number of routers on the attack path, it also holds for any number of distinct markings. Hence, by introducing another way of formulation, we transform the original problem in the case \( f > 1 \) to the formerly solved related problem in the case \( f = 1 \). In other words, **Theorem 1 and Corollary 2 is hold for any value of \( f \).**

Due to the fact that every traceback router divides its IP address into \( f \) fragments of LC bits each to fit the limited bit budget of IPv4 packet header, the marking field of Bit budget bits in IPv4 packet header contains the following four fields: a random coefficient field of \( k \) b-bit coefficient slots \( c_1, \ldots, c_k \), a linear combination field of LC bits to store a linear combination result \( \sum_{i=1}^{f} c_i f_{frag} \) (computed over Galois field \( \mathcal{F}_2^m \)) of \( k \) fragments \( frag \) from some routers on the attack path, a distance field \( dist \) to calculate the distance of the marking router from the collector, and offset field to distinguish drawn fragments with others in the same router.

#### 4.2 Core Network Coding Approach to PPM

##### 4.2.1 Motivation of Core PPM+NC Schemes

In Practical PPM+NC scheme [4] shown in Fig. 3, each traceback router divides its IP address into four 8-bit fragments (i.e., \( f = 4 \)). In this scheme, each marked packet contains a linear combination of \( k = 3 \) fragments with the same offset value \( fID \) from \( k = 3 \) consecutive routers respectively. Therefore, for the bit budget of 17 bits (the 16-bit IP ID field and the 1-bit fragment flag [9]), the marking field in each packet’s IPv4 header contains: a fragment offset field \( fID \) of 2 bits, a random coefficient field of \( k \times b = 6 \) bits, a 1-bit distance field \( dist \), and a linear combination field of 8 bits.

From Theorem 3, we realize that Practical PPM+NC scheme [4] is just one of various ways to design network coding to PPM scheme because we can design PPM+NC schemes so that each marked packet contains a linear combination of \( k \) fragments with any offset values (instead of same values) from any set of routers (instead of consecutive ones). As a result, we propose other schemes called **Core PPM+NC schemes**.

In Appendix, we show that Practical PPM+NC scheme [4] has many limitations. Furthermore, it is unnecessary to encode more than two routers (edge) to reconstruct the attack path. Besides, [2] showed that the further from the victim, the less likely that fragments are perceived. Therefore, we propose Core PPM+NC schemes which focus benefit of network coding in collecting fragments around the marking router rather than distribute such benefit to fragments of the routers nearer to the collector in Practical PPM+NC scheme.

##### 4.2.2 Group Offset

In Practical PPM+NC scheme [4, Fig. 5], each router on the path just contribute only one random fragment of its \( f \) fragments to a linear combination. However, each marked packet received at the victim in Core PPM+NC schemes contains a linear combinations of \( k \) fragments from just one or two routers (marking router or marking router and its next consecutive one) on the attack path, which leads to that each of such routers in Core PPM+NC schemes contributes a random group of its fragments to a linear combination. Therefore, Practical PPM+NC scheme must use \( \log_2 f \) bits for...
fragment offset field in packet’s header to distinguish drawn fragment with other fragments in the same router, whereas Core PPM+NC schemes use less bits for group offset field to distinguish drawn group of fragments with other groups in the same router because each router has \(g\) groups of fragments \((g \leq f)\).

4.2.3 Overlapped Fragments

Instead of using non-overlapped fragments in previous PPM schemes (FMS [2], Practical PPM+NC [4]) or hash fragments in FIT scheme [8], we propose that traceback routers should divide its IP addresses into overlapped fragments (see Fig. 4) to avoid wasteful bits in IPv4 packet header in the case that the remaining bits in bit budget of IPv4 packet header used for linear combination field (or fragment field in other PPM schemes) has more space than the length in bits of each non-overlapped fragment. That is, for a given number of fragments \(f\), each overlapped fragment would have size \(LC \geq \lceil 32/f \rceil\) bits. Note that Practical PPM+NC scheme [4, Fig. 5] uses 4 non-overlapped 8-bit fragments while our Core PPM+NC scheme uses only 13 overlapped 13-bit fragments (Fig. 4).

Using overlapped fragments, allocated space for linear combination field of a marked packet for given \((f, k, b, g)\) in the Core PPM+NC schemes would be:

\[
\begin{align*}
\text{LC} &= \text{Bit budget} - \left\lfloor \log_2 g \right\rfloor - \text{kb} - \text{dist} \\
\frac{f}{\text{LC}} &\geq 32
\end{align*}
\]

4.2.4 Marking Probability

Previous schemes (FMS [2], FIT [8] and Practical PPM+NC [4]) set \(P_m = 1/25\) as the marking probability for every router on the attack path. In order to compare our Core PPM+NC schemes with them in the same coordinate system, we consider our proposed schemes with such setting \((P_m = 1/25)\). Notice that we already proved the benefit of network coding to the PPM in the case of arbitrary marking probabilities in Sect. 3 and Sect. 4.1. Therefore, PPM+NC scheme also requires less expected number of packets to derive the attack path than PPM scheme with setting \(P_m = 1/25\).

4.2.5 Reconstruction Procedure

Each marked packet \(P\) contains: \(\text{random coefficients} = (c_1, \ldots, c_k)\) and \(P_{\text{linear combination}}\) (linear combination result of \(k\) fragments), which corresponds to a linear equa-

\[
\begin{align*}
\text{LC} &= \text{Bit budget} - \left\lfloor \log_2 \left(\frac{2f}{k} \right) \right\rfloor - \text{kb} - \text{dist} - \text{edge} \\
\frac{f}{\text{LC}} &\geq 32
\end{align*}
\]

where \(\text{frag}_s\) are unknowns to the victim. For single-path attack, after collecting \(nf\) innovative marked packets, the victim could infer all \(n\) router IP addresses on the attack path by solving a system of \(nf\) linear equations. For multi-path attack, as mentioned in [4], it is assumed that the victim has a map of upstream routers [7], [8]: For each received marked packet \(P\), based on the value \(P_{\text{dist}}\) in distance field of IPv4 packet header, the victim locates routers at this distance in the built map then it compute linear combinations of possible sub-paths from such candidate routers in order to compare with \(P_{\text{linear combination}}\). If the victim detects a match, it adds the corresponding sub-path to the attack graph.

4.3 Core PPM+NC Scheme I

4.3.1 Packet Marking

In Core PPM+NC scheme I, each marked packet received at the victim contains a linear combination of overlapped fragments from \(\text{two consecutive routers}\) (Fig. 5): the marking router and the next consecutive one. Once a router decides to mark a packet, it draws a random value \(gID\) for group offset field. Then it (the marking router) and the next consecutive one, each of them would contribute a random linear combination of its corresponding group of fragments to the linear combination field. Note that the first coefficient \(c_1\) in Core PPM+NC scheme I is drawn randomly from \(\mathbb{F}_{2^8}\) to avoid collecting no information or zero coefficient vector from the marking router (Sect. 3.2).

The number of groups in each router is \(g = \left\lfloor \frac{2f}{k} \right\rfloor\). If there is remainder of division \(f/(k/2)\) as in Fig. 5, then the last group in every router has less than \(k/2\) fragments. Anyway, number of bits used for group offset field in each marked packet is \(\left\lfloor \log_2 g \right\rfloor\) bits. From (1), we derive the allocated size in bits for linear combination field of every marked packet in the Core PPM+NC scheme I for given \((f, k, b)\):

\[
\begin{align*}
\text{LC} &= \text{Bit budget} - \left\lfloor \log_2 \left(\frac{2f}{k} \right) \right\rfloor - \text{kb} - \text{dist} - \text{edge} \\
\frac{f}{\text{LC}} &\geq 32
\end{align*}
\]
4.3.2 Experimental Results

In our experiment, traceback routers, on a given path of length \( d = 14 \), are set with marking probabilities \( P_m = 1/25 \). As we can see from results of 1000 realizations shown in Fig. 6, our Core PPM+NC scheme I always requires less average number of packets than Practical PPM+NC [4] to derive all routers on a given path with any given \((f, kb)\).

4.4 Core PPM+NC Scheme II

As we know, the advantage of node marking scheme is inferring router at any distance as the victim receives a marked packet [8]. Besides, the limitation of Core PPM+NC scheme I with 1-bit distance mechanism is that it must use more 1 bit edge in order for the next router to determine it is whether the first router after the last marking router or not. Therefore, we propose Core PPM+NC scheme II which combines node marking scheme and network coding.

4.4.1 Packet Marking

In Core PPM+NC scheme II, each marked packet received at the victim contains a linear combination of all \( k = f \) fragments from only the marking router. Interestingly, there are no group offset field (because \( g = 1 \) which leads to \([\log_2 g] = 0\) bit) as in Core PPM+NC scheme I, no fragment offset field as in the previous PPM schemes (FMS [2], AMS [7], FIT [8], Practical PPM+NC [4]) because the random coefficient field of each marked packet in the Core PPM+NC scheme II has both functions: distinguishing fragments and coding. From (1), we derive allocated size in bits for linear combination field of every marked packet in the Core PPM+NC scheme II for given \((f, b)\):

\[
\begin{align*}
\left\lfloor \frac{\text{LC}}{f} \right\rfloor &= \text{Bit budget} - fb - \text{dist} \\
\text{LC} &\geq 32 \\
\end{align*}
\]

Different from Core PPM+NC scheme I, the marking router in the Core PPM+NC scheme II has full right to draw coefficient vector, therefore we propose that the marking router does not distribute the zero coefficient vector to reduce more number of packets required to derive. Furthermore, to lower computation overhead per router for packets marked with the same drawn coefficient vector then same linear combination result, we propose that every router pre-calculates and store a table of \( 2^{k/b} - 1 \) entries \( w_1, \ldots, w_{2^{k/b}-1} \) in its memory, where each entry contains: a non-zero coefficient vector \( w.\text{coefficients} \in \mathbb{R}^b \setminus \{0^b\} \) and its corresponding linear combination result \( w.\text{linear combination} \). Once a router decides to mark the traversing packet, it just need to draw uniformly at random an entry \( w \) from that table: \( w \leftarrow \{w_1, \ldots, w_{2^{k/b}}\} \), and overwrite the marking field of the...
Algorithm 2 Core PPM+NC II: Marking at router

1: for each packet \( P \) do
2: Pick \( u \) uniformly at random from \([0, 1]\)
3: if \( (u < P_m(R)) \) then
4: /*decide to overwrite*/
5: Zero-out all fields
6: \( P_{\text{dist}} \leftarrow \text{TTL}_{[5]} \)
7: \( \text{TTL}_{[4,0]} \leftarrow \text{const} \) \( \text{const} = 10110 \) is optimal value [8]
8: \( w \leftarrow [w_1, \ldots, w_{25}, w_{13}] \)
9: \( \text{Random coefficients} \leftarrow w\text{.coefficients} \)
10: \( \text{Linear combination} \leftarrow w\text{.linear combination} \)
11: else
12: /*decide not to overwrite*/
13: \( \text{TTL} \leftarrow \text{TTL} - 1 \)
14: end if
15: end for

Fig. 7 Core PPM+NC scheme II over given path with \( d = 14 \), \( P_m = 1/25, 1000 \) realizations.

Fig. 8 Marking field in Core PPM+NC scheme II with: \( f = 3 \) overlapped fragments, \( \text{LC} = 13 \) bits, \( c_2 = c_3 = 1 \) bit.

Fig. 9 Experimental results for number of packets needed to reconstruct given paths of varying lengths in Core PPM+NC scheme II and other schemes with parameters: Bit budget = 16 bits, \( P_m = 1/25, 1000 \) realizations.

Packet Marking. For bit budget = 16 bits, every traceback router in our Core PPM+NC scheme II divides its own IP address into 3 overlapped fragments (Fig. 4). Marking field (Fig. 8) of a packet contains: 2-bit random coefficient field (1-bit slot \( c_2 \) for the second fragment, 1-bit slot \( c_3 \) for the last fragment) and 13-bit linear combination field (instead of 2-bit fragment offset field and 13-bit hash fragment field in FIT 4/x scheme [8, Fig. 2]), and 1-bit \( \text{dist} \) distance field.

4.4.2 Experimental Results

As we can see from Fig. 7, our proposed Core PPM+NC scheme II always requires less average number of packets than Practical PPM+NC [4] for any given \((f, kb)\) to derive all routers on the given path with common parameters: \( d = 14 \), \( P_m = 1/25, 1000 \) realizations. For each specific \((f, k)\) (recall that \( k = f \) in Core PPM+NC scheme II), average number of packets is reduced as \( b \) is increased from 1 to 2 bits.

4.4.3 Comparison with Other Schemes

Packet Marking. For bit budget = 16 bits, every traceback router in our Core PPM+NC scheme II divides its own IP address into 3 overlapped fragments (Fig. 4). Marking field (Fig. 8) of a packet contains: 2-bit random coefficient field (1-bit slot \( c_2 \) for the second fragment, 1-bit slot \( c_3 \) for the last fragment) and 13-bit linear combination field (instead of 2-bit fragment offset field and 13-bit hash fragment field in FIT 4/x scheme [8, Fig. 2]), and 1-bit \( \text{dist} \) distance field.

5. Benefit of Network Coding to Coupon Collector’s Problem

Applying network coding to Coupon Collector’s Problem (CCP+NC) is just known to have benefit in the case every linear combinations of \( n \) coupons in the space \( \mathbb{F}_q^n \) over a Galois field has the same probability of receiving as \( 1/q^n \) [5, pp.10–13]. In this section, after revealing the relationship between CCP and PPM in Sect. 5.1 together with the relationship between CCP+NC and PPM+NC in Sect. 5.2, based on our strong theorem for benefit of network coding to PPM in Sect. 3, we prove that there exists numerous distributions over space \( \mathbb{F}_q^n \) so that network coding still offers benefit to CCP in Sect. 5.3. Then we give various designs of network coding to CCP in Sect. 5.4.

5.1 Relationship between PPM and CCP

Theorem 4: Coupon Collector’s Problem could be expressed as collecting \( n \) router IDs in PPM scheme.
Note that $n$ distinct coupons in the CCP have equal probabilities of receiving as $\frac{1}{n}$. In order to collect $n$ router IDs in the PPM scheme, the collector needs to collect $n$ distinct markings, where perceived probability of marking from router ID $i$ is $p_i = P_m(i)$. Therefore, we need to prove that there exists a setting of marking probabilities $(P_m(n), ..., P_m(2), P_m(1))$ for $n$ routers in the PPM scheme so that $p_i = \frac{1}{n}$ $(1 \leq i \leq n)$. For this reason,
\[
\frac{p_i}{p_{i-1}} = \frac{P_m(i) \cdot \prod_{j=1}^{i-1} (1 - P_m(j))}{P_m(i-1) \cdot \prod_{k=1}^{i-2} (1 - P_m(k))} = 1.
\]
\[
\Rightarrow \frac{1}{P_m(i)} = \frac{1}{P_m(i-1)} - 1 \quad (2 \leq i \leq n).
\]
Clearly, the above equation is an arithmetic series. We derive:
\[
\frac{1}{P_m(i)} = \frac{1}{P_m(1)} - i + 1.
\]
Besides, $p_1 = P_m(1) = \frac{1}{n}$ which leads
\[
P_m(i) = \frac{1}{n - i + 1} \quad (1 \leq i \leq n).
\]
As a result, CCP is a special case of PPM scheme where $P_m(i) = \frac{1}{n-i+1} \quad (1 \leq i \leq n)$.

5.2 Abstract Model for CCP+NC and PPM+NC

We recall the scenario of CCP [3, pp.32]: each received box contains one of $n$ distinct coupons $x_1, x_2, ..., x_n$, the coupon in each box is chosen independently and uniformly at random from $n$ possibilities, how many boxes are required to collect every type of coupon? Applying network coding to CCP (CCP+NC) [5, pp.10–13], each box will contains a linear combination of $n$ coupons $\sum_{i=1}^{n} c_i x_i$ instead of an original coupon, where each coefficient $c_i$ is drawn randomly and independently from a Galois field $\mathbb{F}_q$. This transforms the collector’s problem in CCP into collecting $n$ linearly independent vectors from a vector space $\mathbb{F}_q^n$ over a Galois field in order to derive $n$ distinct coupons, which is also the collector’s problem in the PPM+NC scheme (Sect. 3.2). From that analogy between CCP+NC and PPM+NC, we propose an abstract model for CCP+NC and PPM+NC as follows: A source contains all $q^n$ vectors of $n$-dimensional space $\mathbb{F}_q^n$ over the Galois field. Each time, this source randomly distribute a vector $(c_1, c_2, ..., c_n)$ with replacement to the collector. Once obtaining $n$ linearly independent vectors, the collector can get a prize.

However, there is a difference between CCP+NC and PPM+NC: the event of receiving a specific coefficient vector $(c_1, c_2, ..., c_n)$ just depends on events of drawing coefficients $c_i$ in the CCP+NC scheme, while it depends on the event of receiving a marking from a router on the path as well in the PPM+NC scheme. Based on our proposed abstract model for CCP+NC and PPM+NC, that difference between CCP+NC and PPM+NC is interpreted as the difference in probability distributions over $n$-dimensional space $\mathbb{F}_q^n$ at the source.

5.3 Abstract Model of PPM+NC and CCP+NC versus CCP

Theorem 5: There are numerous distributions over a vector space $\mathbb{F}_q^n$ so that collecting problem in the abstract model of PPM+NC and CCP+NC is faster than in CCP.

Let $X_{PPM} (X_{CCP,NC}$ respectively) be the number of packets required to derive $n$ router IDs in the PPM (PPM+NC respectively) scheme with specific setting of marking probabilities for $n$ routers as $P_m(i) = \frac{1}{n-i+1} \quad (1 \leq i \leq n)$. From Theorem 4, we derive:
\[
E[X_{CCP}] = E[X_{PPM}].
\]
Besides, from corollary 2:
\[
E[X_{PPM,NC}] \leq E[X_{PPM}].
\]
Therefore,
\[
E[X_{PPM,NC}] \leq E[X_{CCP}].
\] (4)

In the proof of Theorem 1, it is pointed out that event of receiving a marking of a router at distance $i$ on the attack path in the PPM+NC scheme is the event of receiving a coefficient vector from the set $A_i \subset \mathbb{F}_q^n$ with probability:
\[
p_{A_i} = P_m(i) \cdot \prod_{j=1}^{i-1} (1 - P_m(j)) \quad (1 \leq i \leq n).
\]

With the specific setting of marking probabilities for $n$ routers as $P_m(i) = \frac{1}{n-i+1} \quad (1 \leq i \leq n)$ in the PPM+NC scheme, (4) turns out that a random collecting of $n$ linearly independent vectors from a $n$-dimensional vector space $\mathbb{F}_q^n = A_0 \cup A_{n-1} \cup ... \cup A_1 \cup \{0^n\}$ is faster than CCP provided that $p_{A_i} = \frac{1}{n} \quad (1 \leq i \leq n)$, regardless of $p_{V_A}(v)$ which is the distribution of vectors on set $A_i$. Thus, there are numerous distributions over a vector space $\mathbb{F}_q^n$ so that collecting problem in the abstract model of PPM+NC and CCP+NC is faster than in CCP provided that $p_{A_i} = \frac{1}{n} \quad (1 \leq i \leq n)$. 

5.4 Designing Network Coding to Coupon Collector’s Problem

Theorem 5 motivates us to design network coding to CCP so that probability distribution over $n$-dimensional space $\mathbb{F}_q^n = A_0 \cup A_{n-1} \cup ... \cup A_1 \cup \{0^n\}$ in CCP+NC is: $p_{A_i} = \frac{1}{n} \quad (1 \leq i \leq n)$. We denote $X_{CCP,NC}$ as the number of packets required to collect $n$ linearly independent vectors in such CCP+NC scheme. Then $E[X_{CCP,NC}] = E[X_{PPM+NC}]$, which results in $E[X_{CCP,NC}] \leq E[X_{CCP}]$.

Based on Theorem 5, we design CCP+NC by designing distributions $p_i(k)$ over a Galois field $\mathbb{F}_q$ for coefficients $c_i$ such that $p_{A_i} = \sum_{v \in A_i} p_V(v) = \frac{1}{n} \quad (1 \leq i \leq n)$ as follows:
\[
\Pr(c_i \in F_q \setminus \{0\}) = \sum_{k \in F_q \setminus \{0\}} p_{c_i}(k) = \frac{1}{n+i+1} ,
\]
\[
p_{c_i}(0) = \frac{n-i+1}{n-i+1} , \quad \forall i \leq n-1.
\]
Then,
\[
p_{k} = \Pr(c_i \in F_q \setminus \{0\}). \prod_{1 \leq i \leq k} p_{c_i}(0) = \frac{1}{n} \cdot \frac{n-i+1}{n-i+1} \cdot \frac{n-2}{n-1} \cdot \frac{n-1}{n} \cdot \frac{n}{n} = \frac{1}{n}.
\]

Note that [5, pp.10–13] just compared CCP with the uniform CCP+NC [5, pp.10–13] where \( p_{c_i}(v_j) = \frac{1}{q^r} \) (for \( v_j \in F_q^r \)) due to \( p_{c_i}(k) = \frac{1}{q} \) (for \( k \in F_q \)) for \( 1 \leq i \leq n \). However, we have shown that there are numerous designs of non-uniform distributions over the \( n \)-dimensional vector space \( F_q^n \) so that network coding still offer benefit to CCP provided that those designs satisfy (5).

6. Conclusion

In this paper, we have proposed a novel method to prove the strong theorem for benefit of network coding to probabilistic packet marking in the general case. Based on this theorem, we propose a core network coding approach to PPM (Core PPM+NC schemes) which actually reduce the number of packets required for the victim to infer all routers on the attack path, compared to previous PPM-based schemes. In addition, from the relationship between CCP and PPM combined with our theorem for benefit of network coding to PPM, we prove that there are numerous designs of network coding that still offer benefit to CCP. We hope these findings in this paper give more insight about the power of network coding to probabilistic packet marking and Coupon Collector’s Problem.

References


Appendix: Limitations of Practical PPM+NC Scheme

In this section, we analyse and show drawbacks of Practical PPM+NC scheme [4].

A.1 Unstable Benefit of Network Coding

Practical PPM+NC scheme obtains less benefit from network coding over paths with less than \( k \) traceback routers. In particular, over a path with only one traceback router, Practical PPM+NC is the same as PPM scheme [2].

A.2 Expensive Scheme

It is expensive to keep coding to packet at many routers in flight: each time the marked packet traverses, next traceback routers must access to IP header field for coding.

A.3 Problem of Distance Mechanism

The marking mechanism of Practical PPM+NC scheme described by [4, Algorithm 1] using zero/increment distance mechanism (i.e., 5-bit distance mechanism [2]). The problem arises as Sattari et al. [4] apply the 1-bit distance mechanism (i.e., 5-bit distance mechanism [2]) into Practical PPM+NC scheme. We notice that 1-bit distance mechanism just can calculate actual distance in hops of the marking router, but it cannot count number of traceback routers that a received packet has traversed since the marking router \( R_L \) as 5-bit distance mechanism. According to [4, Alg. 1] with \( k = 3 \) for example, in the case that the next router \( R_{L-1} \) could draw a zero coefficient from \( F_{2^2} \), which makes the next router \( R_{L-2} \) think that it is the first router after the marking router \( R_L \) (because \( R_{L-2} \) checks that the second coefficient slot is empty). Then \( R_{L-2} \) would fill its drawn coefficient into the second coefficient slot. Finally, \( R_{L-3} \) would fill its drawn coefficient into the third coefficient slot. Thus, the the received packet at the victim would contain a linear combination of \( R_L, R_{L-1}, R_{L-2}, R_{L-3} \), but three consecutive routers \( R_L, R_{L-1}, R_{L-2} \) as this scheme expects. Consequently, the victim would infer wrongly three consecutive routers. It means that, the marking mechanism with 1-bit distance field of the Practical PPM+NC scheme in [4] does not work properly.

A.4 More Bits to Count Coding Routers

Because the Practical PPM+NC scheme is broken with 1-bit distance mechanism, it could use more bits to count \( k \) consecutive traceback routers since the marking router. That is, this scheme must use more \( \log_2 k \) bits in the packet’s header.
Dung Tien Ngo received his B.S. degree in information technology at HaNoi University of Technology, VietNam in 2008. Since 2009, he has been working towards the Ph.D. degree at Department of Computer Engineering, Kyung Hee University, South Korea. His research interests include IP Traceback, Randomized Algorithms, Algorithmic Game Theory, Combinatorial Optimization.

Tuan Anh Le received his M.S. and Ph.D. degrees in Computer Engineering from Kyung Hee University, Korea, in 2005, 2012, respectively. He received his BSc degree in Information Technology from Natural Sciences University, Ho Chi Minh city, Vietnam in 2002. His research interests are traffic engineering, flow control, congestion control, and resource management.

Choong Seon Hong received his B.S. and M.S. degrees in electronic engineering from Kyung Hee University, Seoul, Korea, in 1983, 1985, respectively. In 1988 he joined KT, where he worked on Broadband Networks as a member of the technical staff. From September 1993, he joined Keio University, Japan. He received the Ph.D. degree at Keio University in March 1997. He had worked for the Telecommunications Network Lab, KT as a senior member of technical staff and as a director of the networking research team until August 1999. Since September 1999, he has been working as a professor of the School of Electronics and Information, Kyung Hee University. He has served as a Program Committee Member and an Organizing Committee Member for International conferences such as NOMS, IM, APNOMS, E2EMON, CCNC, ADSN, ICPP, DIM, WISA, BeN and TINA. His research interests include ad hoc networks, network security and network management. He is a member of IEEE, IPSJ, KIPS, KICS and KIISE.

Won-Tae Lee received his B.S. and M.S. degrees in electrical engineering from Yonsei University, Korea, in 1983 and 1985, respectively. He received the Ph.D. degree in electronic engineering from Kyungnam University, Korea, in 2006. Since 1985, he has been a principal researcher at Korea Electrotechnology Research Institute (KERI) in Korea. His research interests are smart grid ICT, powerline communication and ubiquitous sensor networks.

Jae-Jo Lee is a senior researcher in Power Telecommunication Network Research Group of KERI (Korea Electrotechnology Research Institute). He received his B.S. degree in physics and M.S. degree in electronic engineering from Kyung Hee University, Seoul, Korea, in 1990, 1992, respectively. In 1992 he joined KERI, where he worked on power telecommunication networks as a researcher. He received the Ph.D. degree in electronics engineering from Kyung Hee University in 2005. From 1999 to 2004, he had worked for the development of power line communication network technology as a senior member of technical staff. Since October 2005, he has worked as a project manager of the ubiquitous power line communication networks development project. His research interests include Power Line Communications, Home Networking and Network Management System. He is a member of Korean Institute of Electrical Engineers (KIEE) and Korean Institute of Communication Sciences (KICS).

Sungwon Lee received the Ph.D. degree from Kyung Hee University, Korea. He is a professor of the Computer Engineering Departments at Kyung Hee University, Korea. Dr. Lee was a senior engineer of Telecommunications and Networks Division at Samsung Electronics Inc. from 1999 to 2008. He is a editor of the Journal of Korean Institute of Information Scientists and Engineers : Computing Practices and Letters.