An Architecture for Network Layer Privacy

Marcelo Bagnulo, Alberto García-Martínez, Arturo Azcorra
Dept. Ingeniería Telemática
U. Carlos III de Madrid (UC3M)
Leganés, Spain
{marcelo, alberto, azcorra}@it.uc3m.es

***Abstract*— We present an architecture for the provision of network layer privacy based on the SHIM6 multihoming protocol. In its basic form, the architecture prevents on-path eavesdroppers from using SHIM6 network layer information to correlate packets that belong to the same communication but use different locators. To achieve this, several extensions to the SHIM6 protocol and to the HBA (Hash Based Addresses) addressing model are defined. On its full-featured mode of operation, hosts can vary dynamically the addresses of the packets of on-going communications. Single-homed hosts can adopt the SHIM6 protocol with the privacy enhancements to benefit from this protection against information collectors.**

Keywords: HBA, IPv6, privacy, SHIM6

I. INTRODUCTION

is a growing concern in today's Internet. This is due to the increasing number of abuses resulting from the malicious use of personal data acquired through Internet. For example, an eavesdropper can track web accesses to disclosure user's interests; or can inspect email exchanges to acquire sensitive personal information. Although an expensive infrastructure is required to correlate IP packets captured in an arbitrary node of the Internet at the application layer, privacy can be compromised by much simpler procedures. For example, the inspection of Internet traffic to obtain characteristics of a given communication such as duration, number of bytes transferred, or parties involved, can provide useful information to a harvester without requiring complex processing. This information can be obtained through proper correlation of identity tokens such as transport layer ports, IP addresses, or link layer addresses in locations close to the communicating nodes. Transport layer information and above can be encrypted by means of IPSec, but even if this costly mechanism is used, network layer identifiers such as IP addresses still provide means to disclose end-point identities and to delimit communication flows.

Additional privacy concerns are raised by the emergence of new protocols enabling mobility (e.g. MIPv6), or multihoming (e.g. SHIM6 [1], currently under development by the IETF). In both cases, the aim is to preserve an established communication when a node changes its network attachment point, i.e. changes the IP address at which it is reachable. The protocol information exchanged among the communicating parties allows an eavesdropper to determine that packets using different IP addresses belong to the same communication, and therefore enables tracking down a given node as it obtains connectivity through different points of the Internet.

In this paper we analyze the privacy concerns raised by the exchange of information performed to maintain SHIM6 communications. In the Basic Privacy mode, the architecture allows conceals sensible information by encrypting parts of the SHIM6 protocol exchange using the secret obtained through Diffie-Hellman. Additionally, the SHIM6-specific identity tokens carried in the data packets to allow the SHIM6 layer to recover the original identifiers are changed for each different path, so that its value is no longer valid for obtaining the multiple IP addresses of a multihomed node, or to gather the pieces of a communication flow when different addresses are used.

Furthermore we can take advantage of the facilities provided by the SHIM6 multihoming protocol to define an architecture that increases the protection against network layer privacy attacks. This additional protection can be provided if the variation over time of the network tokens used by on-going communications is allowed. In this case, the number of packets that can be associated to the same communication by matching network-level identifiers can be reduced at will. This Full-Featured Privacy mode is not only profitable for multi-homed nodes but also for single-homed hosts that may adopt the SHIM6 protocol with the privacy enhancements just to obtain protection from address-based correlation.

In the Full-Featured Privacy mode, the generation and management of the different IPv6 addresses and SHIM6-specific identity tokens relies on pseudorandom sequences that are only known by the communicating parties. The synchronization that is required among the peers is performed without explicit signaling. The management of the pseudorandom sequences of network identity tokens requires the allocation of many tokens (e.g. many IPv6 addresses) per communication. As the number of tokens grows, so does the probability of having an undesirable collision with existent tokens. To assess the validity of the architecture when applied to practical scenarios, a quantitative analysis of the collision of identifiers is performed.

It should be noted that our solution protects from passive eavesdroppers, but does not provide protection against a node that initiates or actively intercepts a communication. Therefore,

This work has been supported by the OPTINET6 project TIC-2003-09042-C03-01 and by the IMPROVISA project TSI2005-07384-C03-02.

1-4244-0353-7/07/$25.00 ©2007 IEEE

an attacker could initiate a communication to obtain a set of addresses assigned to a node, gathering different prefixes to which the node is attached. However, with the Full-Featured Privacy mode, communications are still protected, because the network tokens used are independent from the tokens of any other communication.

The rest of the paper is structured as follows: First the SHIM6 architecture is described, addressing both the protocol and the security model. Next, a privacy analysis to identify the vulnerabilities raised by SHIM6 is presented. The privacy architecture that addresses the vulnerabilities and provides new functionality by extending the protocol is then proposed. After the discussion of the related work and the conclusions, an analysis of the probability of collision of network identity tokens for the Full-Featured Privacy mode is found in the Appendix.

II. SHIM6 OVERVIEW

In order to preserve the scalability of the global routing system, the IPv6 community is limiting the assignment of Provider Independent address blocks to the subscribers of the ISPs, and advocating in that case the adoption of Provider Aggregatable addressing. Such approach forces multihomed sites, i.e. sites connecting to the Internet through multiple providers, to obtain one Provider Aggregatable prefix from each of their provider's address blocks. Moreover, since ISPs only announce their own prefix block into the global routing system, a multihomed host is reachable at a given address only through the corresponding ISP. Consequently, in order to be reachable through all the available ISPs, a host within the multihomed site needs to configure as many addresses as prefixes are available in the site.

The preservation of an establish communication when an outage occurs requires to re-home the communication to the address of an alternative ISP. This change has to be performed transparently with respect to the transport layer, since this layer identifies the endpoints of a communication by the IP addresses of the nodes involved. To manage this change, SHIM6 [1], a multihoming mechanism located within the IP layer, is proposed. The SHIM6 layer performs a mapping between the *identifier* presented to the upper layers and the *locator* actually used to exchange packets on the wire, so different locators could be used to enforce different paths for the communication. To achieve this, two multihomed hosts have to establish a SHIM6 context state in both end-points of the communication so that they can exchange the information about alternative locators using the SHIM6 protocol.

Next we describe the security mechanisms used to protect the SHIM6 architecture, and the architecture of the actual SHIM6 protocol.

A. SHIM6 Security

The ability of the multihoming protocol to associate several locators to a given identifier raises the concern on attacks in which a network identity is redirected to a non-legitimate locator. To protect from redirection attacks when the SHIM6 protocol is used, the use of Hash Based Addresses (HBA) [2] is proposed. HBA are a new type of global IPv6 addresses that incorporate into the interface identifier a cryptographic one-way hash of the prefix-set available in the multihomed host. The result is that the binding between all the addresses of a multihomed host is encoded within the addresses themselves, providing hijacking protection. In a general multihoming scenario, a multihomed host X attached to a link where N 64-bit prefixes are available (PX_1::/64, PX_2::/64,…, PX_N::/64) generates the interface identifier (II) of each of its addresses as a 64-bit hash of the prefix set available in the link and a random nonce (RN):

$$II_J = \text{hash}|_{64} (PX_J::/64,\ PX_1::/64,\ldots,\ PX_{J-1}::/64,\ PX_{J+1}::/64,\ \ldots\ PX_N::/64,\ RN)$$

Finally, the addresses that form the HBA set result from the concatenation of each prefix with its corresponding interface identifier. Note that the interface identifiers are different for each address because the prefixes are placed in different order prior to each hash computation. Any node communicating with a multihomed one can verify through a simple hash calculation that the alternative addresses proposed for diverting the communication are bound to the initial address.

B. SHIM6 Protocol

The SHIM6 protocol [1] creates and manages the SHIM6 context associated with the communication between two nodes, and provides the means to preserve the established communication when data packets using different locators are exchanged. We next describe the context establishment and data packet exchange functions of the SHIM6 protocol.

Consider that one of the nodes involved in a communication decides to create a SHIM6 context in order to benefit from the multihoming capabilities. We refer to the party that initiates the SHIM6 context creation process as the *initiator* and to the other communicating party as the *responder*. We suppose that at least one of the parties is multihomed, in this example the initiator, and its multiple addresses have been generated as an HBA set associated with the multiple prefixes available. The initiator requests the creation of a SHIM6 context associated with a pair of identifiers issuing a message named *I1*. Upon the reception of this message, the responder does not create any state, in order to protect against some forms of Denial-of-Service attacks, but it simply replies with a *R1* message. Then, the initiator sends an *I2* message that contains the following information relevant for us:

- the original pair of identifiers,

- the initiator's context tag, unique for each communication of the initiator, used to identify the data packets that belong to the same communication but contain alternative locators to the original ones, as it is detailed in the next section

- the locator set available at the initiator, and

- the context required to validate the locators at the receiver, e.g. the random nonce required for the HBA generation.

Upon the reception of the *I2* message, the responder verifies that the initiator's identifier is included in the HBA set derived from the received information. If this verification is successful, the responder creates the SHIM6 context, and

replies with a *R2* message, in which it includes its own context tag, locator set and context for address validation.

The SHIM6 layer performs the translation between the identifiers and the locators of the packets send to and received from the wire. Since the identifier is used as a locator until a change occurs, some means to properly identify the packets that need to be translated has to be provided. A context tag (CT) is carried in a *SHIM6 Extension Header* included in the packets that carry addresses other than the identifiers. The CT is uniquely associated to a particular pair of source and destination identifiers. Note that when a packet is sent from the responder to the initiator, the CT set by the initiator is the one that must be included in the packet. Since all the CTs carried in the packets received in the initiator, possibly belonging to different communications, are assigned by the initiator itself, it is easy to assure its uniqueness.

III. SHIM6 PRIVACY ANALYSIS

In this section we analyze the privacy implications of the use of the SHIM6 protocol. The scenario considered consists in two communicating nodes. At least one of them has multiple addresses, so the SHIM6 protocol is used to enable the transparent variation of the addresses used for exchanging data packets. In this scenario, we consider an information harvester H that is placed along the communication path between the two nodes. Depending on the particular location of H, it may be only able to see the packets exchanged using a subset of the addresses available, since different addresses are in general associated to different paths. In the worst case, H is able to see all the packets involved in the communication. In this scenario, H can easily extract useful information about the peers involved in the communication such as number of packets, duration, etc. In addition, H is able to determine some of the multiple addresses that belong to the same host, if he manages to determine part of the address set available for the considered host. We next analyze which information can be extracted by H from the inspection of the SHIM6 protocol.

We consider the case where H is able to inspect the SHIM6 context establishment exchange. During this phase, the peers exchange in clear the identifier pair and the locator set available for each host. If H obtains such information, it can infer that all the addresses contained in each locator set belong to a single host and it can track down its communications, even if the host uses different locators. In addition, during the SHIM6 context establishment handshake, the CTs for each direction of the communication are exchanged. These tags will be included later in data packets carrying alternative locators. An attacker can infer that multiple data packets with different addresses, but carrying the same CT, belong to a given communication, hence discovering that multiple addresses belong to the same host. Note that the CT provides means to correlate packets using different locators even if the context establishment handshake was not captured.

IV. SHIM6 PRIVACY ARCHITECTURE

The architecture proposed consists in a set of extensions to the SHIM6 protocol that can be combined to operate in two different privacy protection modes. In the Basic Privacy mode, a node located along the path will not be able neither to determine the HBA address sets of the multihomed nodes, nor to correlate different packets exchanged through alternative paths. To achieve this, we conceal the relevant information exchanged in the SHIM6 protocol (locator sets, address validation parameters, and context tags), and we enforce the use of different CTs for packets that carry different addresses but belong to the same communication.

In the Full-Featured Privacy mode, besides the protection of the SHIM6 exchange, the dynamic variation of locators and CTs for an on-going communication is enabled. Sequences of interface identifiers and CTs are defined, so that they can be changed dynamically for a given path. If each variation of the interface identifiers and the CT is combined with a change of the prefix pair used (avoiding previously used <local prefix, remote prefix> pairs), the difficulty in using network tokens for the correlation of chunks of transferred data is substantially increased. Moreover, this mode can be used to provide some kind of privacy even for a single-homed host, since it can vary its addresses over time for an on-going communication although the transmission path remained unchanged.

It is worth to note that the variation over time of the CTs or addresses may result in collisions in a host, so a mechanism to handle gracefully these collisions should be provided. A *CT collision* occurs when a host allocates the same CT for two communications, defeating the SHIM6 flow identification capacity. An *address collision* occurs when the dynamic address generation process results in an address that is already in use by another node in the segment in which it is defined, and it may result in the disruption of the communication.

We next detail the mechanisms common to both privacy modes, namely a modification in the SHIM6 context establishment phase to encrypt the sensible information carried in the SHIM6 protocol with a secret agreed through a Diffie-Hellman exchange, the negotiation of the use of privacy facilities for a communication, and a simple negotiation of the privacy mode. Then we briefly address how the Basic Privacy mode provides different CTs for each prefix assigned to the node. We finish with the description of the mechanisms introduced by the Full-Featured Privacy mode to dynamically vary addresses and CTs.

A. SHIM6 Protocol Exchange Extensions

In order to conceal the information exchanged in the SHIM6 context establishment exchange, we propose the use of the Diffie-Hellman key agreement to generate a shared secret, and the use of this shared secret to encrypt the information that must remain private. The Diffie-Hellman key agreement is included as an option in messages *R1* and *I2*, so that after the reception of *R1*, the initiator can create the shared secret and conceal the private information (the locator set, address validation parameters, and the parameters required for the generation of the sequences of CTs and addresses, as shown below) included in the *I2* packet. Analogously, *R2* can be used by the responder to include its corresponding information.

The negotiation for the extended privacy facilities occurs as follows: The request for privacy can be raised by either the initiator or by the responder. The initiator can include a *Privacy*

Request option in the *I1* message to notify the responder its desire to obtain privacy support for the communication for which the SHIM6 context is being created. Either as a response to this request, or as a consequence of the responder's own will, the responder can initiate the Diffie-Hellman exchange at message *R1*. If either the responder or the initiator does not exchange the cryptographic material required to generate the Diffie-Hellman key, the communication continues without privacy support. Since unknown SHIM6 options are silently discarded, a communication with a node that does not implement the privacy extensions seamlessly fall back to the standard SHIM6 protocol exchange.

To negotiate the Privacy mode used for a communication, a *Privacy Mode* option is defined with two possible values corresponding to the Basic and to the Full-Featured Privacy modes. This option can be included in both *I2* and *R2* messages to allow each peer to express the configuration desired. The resulting configuration is the Basic Privacy mode unless both nodes request the Full-Featured Privacy mode.

B. Context Tag Management for Basic Privacy Mode

In the Basic Privacy mode, locators and CTs are stable for the duration of the communication. However, a modification to the current SHIM6 specification is required in order to enforce the use of different CTs per possible locator pair of the communication, so an eavesdropper could no longer use the CTs to correlate information. Both parties must agree in the CTs to be used in order to be able to recover the original identifiers.

The CT generation process is as follows: Both node X (with prefixes PX_1, PX_2, …, PX_N) and node Y (with prefixes PY_1, PY_2, …, PY_M) generate randomly a seed that is unique for the communication, respectively seedX and seedY. The CT at node X, for a given communication between node X and Y, associated to the locator pair $<PX_J , PY_K>$ (i.e. the CT that would be included in data packets sent from locator PY_K to locator PX_J), is computed as:

$$CT(X)_{J\,K} = hash|_{47}(PX_J, K, seedX)$$

Consequently, its corresponding CT at node Y is

$$CT(Y)_{K\,J} = hash|_{47}(PY_K, J, seedY)$$

Both nodes can compute all the CTs at X and at Y. Since the prefixes are encrypted for its transmission in the SHIM6 context establishment exchange, an eavesdropper cannot determine the CTs used in the communication.

A collision may occur in each node if any of the local CTs generated for this communication has been previously allocated for any other SHIM6 communication. To avoid a collision, the node should check that there are no coincidences between the CTs candidate to be allocated and the CTs of prior active communications. If a coincidence occurs, new seeds are generated and tested until no collision is found. If we compute the probability in a similar way to the one presented in the Appendix, for a restrictive case in which the nodes hold 100 private communications, and 100 prefixes each, the probability of collision is still below $7.1 \cdot 10^{-5}$.

In order to avoid the need for additional messages the seeds should be transmitted in messages *I2* and *R2* of the SHIM6 context establishment exchange. However, the initiator does not know neither the locator set of the receiver nor the number of prefixes of the receiver until *R2* is received. It can test a seed assigning to the number of prefixes of the receiver a number large enough so that the real number would not exceed it. After receiving the *R2* message, it can discard the extra CTs. If the number of prefixes received in *R2* from the remote node would be larger than the number allocated for the collision test, new checks are performed for the untested CTs. If the test fails, new seeds are obtained until successfully checked for the real number of prefixes of the receiver, and a SHIM6 Update Request message is sent by the initiator.

C. Dynamic Management of Locators and Context Tags for the Full-Featured Privacy Mode

In this section we propose the use of keyed pseudorandom sequences for the generation of the addresses and CTs, so that both nodes, and only these nodes, can generate the next values that are to be used when a variation occurs.

First of all we want to state that all tokens must change jointly when a variation is raised, because otherwise any unchanged token could be used by an attacker to defeat the protection. Therefore after a *privacy instance* p defined by

$$<locator(X)^p, CT(X)^p, locator(Y)^p, CT(Y)^p>,$$

a new privacy instance p+1 with

$$<locator(X)^{p+1}, CT(X)^{p+1}, locator(Y)^{p+1}, CT(Y)^{p+1}>$$

is used.

The pseudorandom sequences are built upon a secret seed defined by each node, and upon a shared *privacy instance index*. The seed is encrypted with the key generated by the Diffie-Hellman exchange, and conveyed to the correspondent peer in messages *I2* and *R2*. The privacy instance index is initialized to 0 in both nodes, and it is increased by 1 each time a new set of local and remote addresses and CTs are selected.

To enable the dynamic generation of addresses, we define a new *Privacy Extension* that is appended to the HBA structure, so that it is included in the input to the hash that generates the interface identifier for each address. When the privacy instance index is 0, i.e. before any change in the locators occurs, the valid locators are the ones obtained from the HBA without including the Privacy Extension. For privacy instance indexes above 0, the interface identifiers in the private context p corresponding to prefix PX_J are computed as follows:

$$II_J^p = hash|_{64}(PX_J::/64, PX_1::/64, \ldots, PX_{J-1}::/64, PX_{J+1}::/64, \ldots, PX_N::/64, RN, seedX, p)$$

In the Full-Featured Privacy mode, CTs are not required to be different for each pair of locators in the same privacy instance index, since only one pair of locators will be used for a given privacy instance index. Therefore, CTs can be computed as:

$$CT(X)^p = hash|_{47}(seedX, p)$$

The CT at node X is

$CT(Y)^p = hash|_{47}$ (seedY, p)

The avoidance of collisions in the Full-Featured Privacy mode is based on the *a priori* reservation of a collision-free set of sequential privacy instances (including Interface Identifiers and CTs) for each private communication. A node that may initiate or receive a communication with SHIM6 privacy generates a local seed and defines a number of consecutive instances for which collisions should be avoided. Then, the node generates the local CTs for each of the contexts to be specified and checks for a CT collision. If no collision occurs, then all the local addresses corresponding to the contexts specified are generated, and configured in its corresponding interfaces, so that the IPv6 Duplicate Address Detection procedure is started for each one to check for address collisions. The Duplicate Address Detection procedure can be performed in parallel for all the candidate addresses, so that the delay introduced, a default value of 1 second, did not depend on the number of addresses. Note that the same interface identifier must be unique only in the segment to which it is assigned. If no duplicate addresses are found, then the set of instances are valid for its use in the near future. If either a CT or address collision is detected, an alternative local seed is generated and the process is restarted. The probability of any collision, which is the probability of being forced to repeat the privacy instance allocation, is very low, less than 10^{-5} even for scenarios with intensive use of privacy facilities, as it is shown in the Appendix.

The number of consecutive instances allocated for the communication by each host is exchanged in *I2* and *R2* of the SHIM6 context establishment phase. Each host can select a different number of consecutive collision-free instances, and the minimum value is selected by the peers. The instances can be used in a round robin fashion, although it is highly desirable to avoid the use of the same instance twice.

Once the communication is established, the synchronization is coordinated as follows: a given privacy instance is used until one of the communicating peers decides to change it, either because of the expiration of a timer, that could be managed by the operating system, or due to an instance change requested by an application. When a node receives a packet with different addresses and local CT, it changes the current instance to the one corresponding to the received packet. Note that this procedure does not require any explicit signaling that could provide hints to malicious information collectors.

V. Related Work

RFC3041 [3] proposes the generation of different interface identifiers for IPv6 address in order to provide protection against the correlation of different communications. However, it does not provide any means to use different addresses for the same communication, and it does not consider the problems raised by the use of multihoming.

The use of pseudo random sequences as a privacy enhancement for network and transport layers is discussed in [4], from which some concepts of this paper have been drawn, although they do not consider multihoming neither as a problem (if the exchange protocol is not properly adapted) nor as an enabler for privacy protection.

The MIPv6 protocol shares similar privacy concerns with multihoming. In the Route Optimization mode, the Home Address (HoA) is carried on each data packet so it can be used to trace different locations of a Mobile Node. In [5] the HoA is substituted by a Privacy Label that is generated pseudo randomly from information exchanged using the protected path through the Home Agent. Although it is stated that the privacy label could vary even in each packet, there is no description on how this could be synchronized, the effect of packet loss in the sequencing, how collisions are treated, as it is discussed here.

Other proposals rely on middleboxes to provide privacy. Application layer proxies are used to encrypt the data transferred through the Onion Routing overlay network [6]. Castellucia et al. [7] propose an authorization model based on cryptographic identifiers to establish IPsec tunnels between specific security gateways. In both cases, network layer tokens are protected only in the path between intermediate nodes, protection depends on confidence on third parties, and costly infrastructure is required.

As mentioned before, our solution does not provide protection against an eavesdropper that initiates or actively intercepts a communication. In the BLIND framework [8], the end-point identifier is based on asymmetric cryptography and completely decoupled from the IP locators. A Diffie-Hellman key exchange protocol protects the identifier even from active attackers. However, this holds for a limited scenario, since *a priori* knowledge of the identifiers of the participants is required, while the identifiers must not be kwon by the attackers. In the SHIM6 multihoming architecture, the identifiers are also valid locators to allow easy deployment when IP addresses are passed as referrals among applications.

VI. Conclusions

We have presented a SHIM6-based architecture for the provision of network layer privacy for multihomed hosts that can be also be used to provide privacy to single-homed hosts. The architecture defines some extensions to the SHIM6 protocol and to the HBA address generation process that allow the provision of different levels of privacy. In the Basic Privacy mode, a multihomed host conceals the information exchanged by the SHIM6 protocol by means of a Diffie-Hellman key agreement option, and the data packets use different context tags for each path. If the dynamic variation of network tokens is desired, the Full-Featured Privacy mode allows the management of a sequence of context tags and interface identifiers that change over time for an on-going communication, without explicit signaling. A single-homed node can take advantage of the possibility of using addresses that change over time even if the communication path is not changed. In this case, several different prefixes should be assigned to each segment if the address space is large enough to difficult address correlation.

This architecture introduces some costs compared to the standard SHIM6 one. First, new context state is required at the communicating peers: more context tags and addresses per prefix, seeds, privacy instance index, and number of instances allocated. Computing costs are incurred for collision checks, and eventually, with a very low probability, for obtaining new

seeds in case of collision. Another cost component is derived from the CPU time required for the Diffie-Hellman key agreement and the encryption of the relevant privacy parameters. However, it should be highlighted that no burden is placed on data packets. Future work, quantitative analysis of the performance impact should be conducted.

The resulting architecture provides by itself a certain degree of privacy, since it allows a node to use different addresses during the lifetime of a communication making it harder to track an ongoing communication. However, in order to achieve full privacy protection, the architecture should be combined with additional mechanisms to conceal higher layer information. It is also required to conceal lower layer information if the eavesdropper shares the link layer with one of the communication peers. For future work, protection to IPsec parameters should be provided by the SHIM6 privacy architecture, so that identity tokens such as the SPI index could not be used for packet correlation.

REFERENCES

[1] E. Nordmark and M. Bagnulo, "Level 3 multihoming shim protocol" IETF Internet-Draft draft-ietf-shim6-proto-07.txt (work in progress), November 2006.

[2] M. Bagnulo, A. Garcia-Martinez and A. Azcorra, "Efficient Security for IPv6 Multihoming", ACM Computer Communications Review, Vol. 35, n. 2, ACM Press, pp. 61-68, April 2005 .

[3] T. Narten and R. Draves, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", IETF RFC 3041, January 2001.

[4] J. Arkko, P. Nikander and M. Näslund, "Enhancing Privacy with Shared Pseudo Random Sequences". 13th International Workshop on Security Protocols, Cambridge, 2005.

[5] R. Koodli, V. Devarapalli, H. Flinck and C. Perkins, "Solutions for IP Address Location Privacy in the presence of IP Mobility". IETF Internet-Draft, draft-koodli-mip6-location-privacy-solutions-00.txt (work in progress), 2005.

[6] M. Reed, P. Syverson and D. Goldschlag, "Anonymous connections and Onion Routing". IEEE J. Selected Areas in Communications 16, 4, pp. 482–494, May 1998.

[7] C. Castellucia, G. Montenegro, J. Laganier and C. Neumann, "Hindering Eavesdropping via IPv6 Opportunistic Encryption". ESORICS 2004, LNCS 3193, pp. 309-321, 2004.

[8] J. Ylitalo and P. Nikander, "BLIND: A Complete Identity Protection Framework for End-points". International Workshop on Security Protocols, Cambridge, 2004.

APPENDIX: PRIVACY INSTANCE COLLISION PROBABILITY IN THE FULL-FEATURED PRIVACY MODE

The solution for this problem is a variation of the well known "birthday problem". Suppose an existing group of j numbers for which there are no collisions, and k numbers are added to the group. The k added numbers are the result of an integer random variable, with uniform distribution between 1 and n. Then, the number of ways we can choose k values out of n, without generating duplicates in the $j+k$ numbers of the new group is $(n-j) \cdot (n-j-1) \cdot \ldots \cdot (n-j-k+1)$.

On the other hand, the number of possibilities for choosing k elements out of n without the restriction of not having any duplicates is n^k. Therefore, the probability of having at least a collision when we select k elements out of n and we add them to a group with j elements is

$$P(n,j,k) = 1 - \frac{(n-j)\cdot(n-j-1)\cdot\ldots\cdot(n-j-k+1)}{n^k} \quad (1)$$

Performing simple computations in (1) we obtain

$$P(n,j,k) = 1 - \left\{\left(1-\frac{j}{n}\right)\cdot\left(1-\frac{j+1}{n}\right)\cdot\ldots\cdot\left(1-\frac{j+k-1}{n}\right)\right\} \quad (2)$$

Since $\forall i \in [1,2,\ldots,k-1] \Rightarrow \frac{i}{n} \le \frac{k-1}{n}$, and $k < n$, we can bound $P(n,j,k)$ to avoid the computation of factorial numbers

$$P(n,j,k) \le 1 - \left(1-\frac{j+k-1}{n}\right)^{k-1} = B(n,j,k) \quad (3)$$

To particularize the formula to the Full-Featured Operation Mode, we define the following parameters:

C – Number of communications per node using the privacy extensions

W – Privacy window size per communication context

N – Number of nodes per IPv6 segment

X – Number of prefixes per node

Another data are the size of the CT space, 2^{47}, and the number of different HBA interface identifiers in a segment, 2^{59} (five bits of the interface identifier are reserved for other uses, although we have omitted the details for the sake of clarity).

Combining the probabilities of CT collision in the node, and address collision on each prefix allocated to the node, we can obtain an upper bound for the probability of a privacy instance collision (P_{PIC}):

$$P_{PIC} \le 1 - \left[1 - B(2^{47}, W\cdot C, W)\right]\left[1 - B(2^{59}, W\cdot C\cdot N, W\cdot C)\right]^X \quad (4)$$

For C=10, W=100, N=100, X=10, the probability of collision is less than $2.5\cdot10^{-9}$. Even for intensive use of the privacy facilities, in which each segment is shared among a large number of systems (increasing the possibility of an address collision), C=100, W=1000, N=1000, X=10, the probability of collision is less than $1.8\cdot10^{-5}$.