

Анонимная сеть «Hidden Lake»

Коваленко Геннадий Александрович

Аннотация. Сеть Hidden Lake, являясь по природе своей QВ-сетью, представляет собой также ряд новых архитектурных решений, ранее не применявшихся в строении анонимных систем. Базируясь на принципе микросервисной архитектуры, таковая сеть может не только добавлять, но также и удалять функции по мере своей необходимости, никак не изменяя при этом общий механизм работы. Базируясь на слепой маршрутизации и полном шифровании сообщений, таковая сеть связывает всех узлов в системе, не позволяя применять долговременное наблюдение за связями и фактом коммуникации. Понимание общих принципов работы сети на базе её математических моделей способно дать не только оценку корректности функционирования всей системы, но и также возможный вектор развития будущих анонимных коммуникаций.

Ключевые слова: скрытые системы; анонимные сети; децентрализованные сети; теоретически доказуемая анонимность; qb-задача; микросервисная архитектура; сеть hidden lake;

Содержание

1. Введение.....	1
2. QВ-задача.....	2
2.1. Недостатки QВ-сетей.....	8
2.2. Активные наблюдения.....	9
2.3. Сравнение с другими задачами.....	11
3. Функция шифрования.....	12
3.1. Первый этап шифрования.....	13
3.2. Второй этап шифрования.....	14
4. Сетевое взаимодействие.....	15
4.1. Микросервисная архитектура.....	15
4.2. Стек протоколов «GP/12».....	17
5. Структурные параметры.....	19
6. Заключение.....	24

1. Введение

Анонимная сеть Hidden Lake (HL) - это децентрализованная F2F (friend-to-friend) [1] анонимная сеть с теоретической доказуемостью [2, с.49]. В отличие от известных анонимных сетей, подобия Tor, I2P, Mixminion, Crowds и т.п., сеть HL способна противостоять атакам глобального наблюдателя. Сети Hidden Lake для анонимизации своего трафика не важны

такие критерии как: 1) уровень сетевой централизации, 2) количество узлов, 3) расположение узлов и 4) связь между узлами в сети, что делает таковую систему абстрактной [2, с.144].

2. QB-задача

Задача на базе очередей (QB - Queue Based) [2 с.149] представляет собой ядро анонимной сети Hidden Lake за счёт которого формируется теоретически доказуемая анонимность. QB-сети представляют собой одну из наиболее простых задач анонимизации в плане программной реализации, в сравнении с другими представителями теоретической доказуемости в лице DC (Dining Cryptographers) [3, с.225] и EI (Entropy Increase) [2, с.165] - сетей. Формально QB-сеть можно описать системой следующего вида:

$$QB-net = \Sigma_{i=1}^n (T = \{t_i\}, K = \{k_i\}, C = \{(c \in \{E_{k_j}(m), E_{r_i}(v)\}) \leftarrow^{t_i} Q_i\})$$

где n - количество узлов в системе, K - множество ключей шифрования, T - множество периодов генерации, C - множество зашифрованных сообщений, Q - очередь зашифрованных сообщений, i, j - идентификаторы отдельных узлов, E - функция шифрования, m - открытое сообщение, v - ложное сообщение, r - ключ шифрования не находящийся во множестве K .

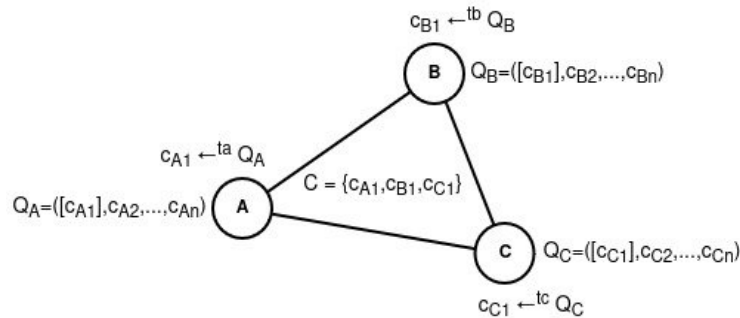


Рисунок 1. QB-сеть с тремя участниками A, B, C

Вышеуказанная система может быть представлена четырьмя состояниями:

1. $Q_i \leftarrow (c = E_{k_j}(m))$, где $k_j \in K$, $c \in C$. Открытое сообщение m шифруется ключом получателя $k_j \in K$. Результат шифрования $c = E_{k_j}(m)$ помещается в очередь Q_i ,
2. $(c = E_{k_j}(m)) \leftarrow^t Q_i$, если $Q_i \neq \emptyset$, где $t \in T$, $k_j \in K$, $c \in C$. В каждый период времени t из очереди Q берётся зашифрованное сообщение c и отправляется всем участникам сети,
3. $(c = E_{r_i}(v)) \leftarrow^t Q_i$, если $Q_i = \emptyset$, где $t \in T$, $r_i \notin K$, $c \in C$. Если на период времени t очередь Q остаётся пустой, то создаётся ложное сообщение v , которое далее шифруется ключом без получателя $r_i \notin K$. Результат шифрования $c = E_{r_i}(v)$ отправляется всем участникам сети,
4. $m' = D_{k_i}^{-1}(c)$, где $c \in C$. Каждый участник пытается расшифровать полученное зашифрованное сообщение c из сети своим ключом k_i^{-1} . Если сообщение не поддаётся расшифрованию $m' \neq m$, то это значит, что получателем является либо кто-то другой (использован ключ $k_j \in K$), либо никто (использован ключ $r_j \notin K$).

Теорема 1. В множестве ключей $r \in R$, для любого шифртекста $c = E_r(v)$, существует такое r^{-1} , которое приводит к расшифрованию ложного сообщения $v = D_{r^{-1}}(c)$.

Доказательство 1. Множество R определяется разностью двух множеств $U \setminus K$, при $U = K \cup R$, представляющим собой множество ключей функции шифрования. Для любого $u \in U$ выполняется отображение в множество шифртекстов: $\bigcup E_u \rightarrow C$, при котором также будет существовать такое u^{-1} , выполняющее обратное отображение в множество открытых текстов: $\bigcup D_u^{-1} \rightarrow M$, исходя из того, что система $\Sigma(M, C, U, E, D)$ является шифром [4, с.75].

Анонимность QВ-сетей базируется на сложности определения состояния шифрованного сообщения c , а именно чем оно является: $E_{ki}(m)$ или $E_r(v)$. Очередность сообщений Q в свою очередь гарантирует, что всегда будет существовать такое сообщение c , которое будет сгенерировано системой в период времени равный t , независимо от природы самого сообщения. При отсутствии истинных сообщений очередь Q можно рассматривать как очередь исключительно ложных сообщений $E_r(v)$. При появлении истинного сообщения $E_{ki}(m)$, таковое начинает заменять собой ложное $E_r(v)$ в определённый период времени t . Вследствие этого, QВ-сеть становится генератором шума с функцией кратковременной замены случайного трафика на действительный, а неразличимость шифрованных сообщений друг от друга становится ключевым фактором анонимности.

Теорема 2. При наличии двух ключей шифрования k, r (порождающих соответственно истинные и ложные шифртексты), определение истинности выбранного шифртекста c_i из конечного множества $C = \{c_1, c_2, \dots, c_n\}$ сводится к вычислительно сложной задаче, если соблюдаются условия безопасности E, k, r параметров.

Доказательство 2. Задача истинности шифртекста c_i из множества C сводится к определению его принадлежности к двум состояниям: $E_k(m_i)$ и $E_r(v_i)$. При неизвестных параметрах k, r верхняя граница поиска определяется количеством итераций полного перебора равным $|K \cup R| = |U| = |\{u_1, u_2, \dots, u_{|U|}\}|$, при $K \cap R = \emptyset$ соответственно, т.к. неизвестной переменной в данном случае остаётся принадлежность перебираемых значений u_i к множествам K и R (Теорема 1). При известном u , но неизвестном u^{-1} соответственно, проблема определения истинности шифртекста c_i сводится к одной из асимметричных задач [3, с.378][3, с.386].

Итого, анонимность QВ-сетей определяется не только разрывом связи между отправителем и получателем для глобального наблюдателя, но и также отсутствием связи в самом факте отправления и получения информации. Иными словами, для пассивных наблюдателей, включающих в себя и глобального наблюдателя, ставится непосильной задача определения состояния субъекта, а именно:

1. Отправляет ли участник i в период равный t_i истинное сообщение $E_{ki}(m)$?
2. Получает ли участник i в периоды равные $T \setminus \{t_i\}$ какое-либо сообщение $D_{ki}^{-1}(c)$?
3. Бездействует ли участник $i \rightarrow E_r(v)$ в анализируемом периоде t_i ?

При всех таких сценариях, не будучи одним из узлов участвующих непосредственно в коммуникации и не проявляющим какое-либо влияние на очередь Q_i анализируемого участника i , т.е. не будучи узлом проявляющим активное наблюдение, задача считается невыполнимой, если алгоритм шифрования E и ключи k, r являются надёжными.

Алгоритм 1. Функционирование участника i в системе $QB-net$ на языке псевдокода

ВВОД: очередь Q_i , период t_i , функции E, D , ключи шифрования k_i, r_i

ВЫВОД: подмножество шифртекстов $C_i \in C$

thread-1. (* Генерация истинных шифртекстов *)

for (;;) {

$k_j \leftarrow INPUT_STREAM$ (* Ввод ключа получателя *)

$m \leftarrow INPUT_STREAM$ (* Ввод открытого текста *)

$Q_i \leftarrow (c = E_{k_j}(m))$

}

thread-2. (* Генерация ложных шифртекстов *)

for (;;) {

 if ($Q_i = \emptyset$) {

$v \leftarrow RANDOM_STREAM$ (* Получение ложного текста от КСГПСЧ *)

$Q_i \leftarrow (c = E_{r_i}(v))$

 }

}

thread-3. (* Отправление шифртекстов в сеть *)

for ($a = 1$; ; $a = a + 1$) {

$sleep(t_i)$ (* Ожидание периода *)

$c_a \leftarrow Q_i$, где $c_a \in C_i$

$QB-net \leftarrow c_a$ (* Запись шифртекста в сеть *)

}

thread-4. (* Принятие шифртекстов из сети *)

for ($b = 1$; ; $b = b + 1$) {

$c_b \leftarrow QB-net$, где $c_b \in C \setminus C_i$ (* Чтение шифртекста из сети *)

 if $valid(m = D_{k_i}^{-1}(c_b))$ { (* Проверка корректности расшифрования *)

$OUTPUT_STREAM \leftarrow m$ (* Вывод полученного открытого текста *)

 }

}

В QB -сетях есть также ряд интересных и не совсем очевидных моментов. Так например, период t_i каждого отдельного участника i не обязательно должен иметь константное значение. Период может изменяться по времени или вовсе иметь случайное значение. Такое поведение никак не отразится на анонимности узлов до тех пор, пока будет существовать сам факт отложенности сообщений s в лице очереди сообщений Q . Если сообщение можно будет отправлять в обход очередности, тогда анонимность будет постепенно ухудшаться в зависимости от количества отправляемых подобным способом сообщений.

Таким образом, в отличие от DC-сетей, где период T представлен только одним общим значением $T \setminus \{t\} = \emptyset$, QB-сети делают период не только субъективно (индивидуально) настраиваемым $T = \{t_1, t_2, \dots, t_n\}$, но также и не обязательно статичным для каждого генерируемого сообщения $t \in [l;k]$, где $l \leq k$. Такое свойство позволяет QB-сетям не кооперировать с отдельными узлами за период, а также более качественно скрывать закономерность принадлежности пользователя к анонимизирующему трафику.

Теорема 3. Теоретически доказуемая анонимность *QB-net* системы основывается на систематичности порождения множества шифртекстов $C = C^r \cup C^k = \{c_1, c_2, \dots, c_n\}$ и на неразличимости его подмножеств относительно истинных C^k и ложных C^r шифртекстов.

Доказательство 3. Пусть $C^r = \{c_1^r, c_2^r, \dots, c_m^r\}$ есть множество ложных шифртекстов, а $C^k = \{c_1^k, c_2^k, \dots, c_n^k\}$ напротив есть множество истинных шифртекстов при $m, n \geq 0$ соответственно, тогда при их объединении создаётся множество всех шифртекстов: $C^r \cup C^k = \{c_1^r, c_2^r, \dots, c_m^r, c_1^k, c_2^k, \dots, c_n^k\} = \{c_1, c_2, \dots, c_{m+n}\} = C$. Пусть под задачей деанонимизации $P_{(d)}$ будет далее пониматься однозначное нахождение факта существования или отсутствия истинного сообщения ($n > 0$?) в множестве шифртекстов C . Задача деанонимизации, в свою очередь, опирается на две другие задачи: неразличимости $P_{(i)}$ и систематичности $P_{(s)}$. При решении одной из двух подзадач, задача деанонимизации будет считаться также выполнимой, что можно выразить дизъюнктивной формой: $P_{(d)} = P_{(i)} \vee P_{(s)}$.

Задача неразличимости $P_{(i)}$ может быть определена двумя возможными ситуациями: 1) либо способом нахождения истинного шифртекста $c_i = E_k(m_i) \in C^k$; 2) либо, напротив, способом доказательства отсутствия истинных шифртекстов $c_i \notin C^k = \emptyset$. Задача становится тривиальной при отсутствии шифртекстов в общем, т.к. $C = \emptyset \rightarrow C^k = \emptyset$. Если же $C \neq \emptyset$, тогда задача сводится к проблеме соотношения шифртекстов $c_i \in C$ к их первоначальным подмножествам: C^r или C^k , что, как было показано ранее, является вычислительно сложной задачей (Теорема 2).

Задача систематичности $P_{(s)}$ может быть определена нахождением дополнительных связей в механизме генерации шифртекстов $C = \{c_1, c_2, \dots, c_n\}$. Если взять частный случай $C = \emptyset$, то связность шифртекстов будет априори отсутствовать, что, в свою очередь, показывает вводное отличие задачи систематичности от неразличимости. Далее, пусть $|C| = 1$, т.е. $C = \{c\}$, тогда, задача систематичности будет сводиться к вопросу: «вследствие какого события x был получен шифртекст c ?». Если событие x не имеет связей с каким-либо открытым текстом m (полученным, получаемым, отправленным или отправляемым), т.е. шифртекст c не был создан вследствие появления m как события, тогда x следует рассматривать как *независимое* событие. Если предположить, что существует некий алгоритм A , генерирующий шифртексты c_i посредством независимого события x , т.е. $c_i \leftarrow A(x)$, где $c_i \in C$, и при этом $C = C^r$, тогда в механизме генерации априори будут отсутствовать дополнительные связи кроме основной связи в лице события x , т.к. $c_i \notin C^k = \emptyset$. Теперь, если предположить обратное: $C = C^r \cup C^k$, где $C^k \neq \emptyset$, то сохранение единой связи генерации, в лице независимого события x , становится возможным тогда и только тогда, когда шифртексты $c_i \in C^k$ будут создаваться вместе и вследствие того же алгоритма A , что и шифртексты подмножества C^r , посредством исполнения условия: $(c = E_k(m)) \text{ if } (m \neq \text{null}) \text{ else } (c = E_r(v)) \leftarrow A(x)$. В результате этого, истинная связь ($m \neq \text{null}$) инкапсулируется в независимой связи x , и не проявляет, тем самым, свою природу генерации.

Следствие 3.1. Конечное множество шифртекстов C можно рассматривать как результат поэтапной генерации подмножеств $C_1 = \{c_1\}$, $C_2 = \{c_1, c_2\}$, ..., $C_n = \{c_1, c_2, \dots, c_n\}$ и как завершение сетевой коммуникации в общем. Таким образом, при выполнении условий неразличимости и систематичности множеством $C = C_n$, подмножества $C_i \subseteq C_n$ продолжают в равной степени наследовать их выполнение.

Следствие 3.2. Пусть t есть период генерации порождаемых подмножеств C_i . В таком случае, t есть также вводное условие и независимое событие для алгоритма генерации шифртекстов: $c_i \leftarrow A(t)$. Раз период t базируется на алгоритме A и при этом порождение шифртекстов не связано с открытыми текстами, тогда его варьируемая характеристика в лице статичности ($t = T$) или динамичности ($t \in [l;k]$) не способна воздействовать на качество анонимности.

Следствие 3.3. Теоретически доказуемая анонимность QВ-задачи не зависит от какого-либо одного конкретного алгоритма генерации шифртекстов A , и как следствие, множество участников системы $\{1, 2, \dots, n\}$ способно использовать множество различных алгоритмов $\{A_1, A_2, \dots, A_n\}$, по причине достаточности условия в лице независимости событий при генерации шифртекстов.

Пример 3.1. В QВ-задаче под алгоритмом A наиболее часто понимается периодичность генерации шифртекстов на базе очередей с вводным условием t (периодом). Хотя это и наиболее практичный алгоритм, он всё же не единственен. Так например, под алгоритмом генерации A' может пониматься также формирование шифртекста $c_i \leftarrow A'(n)$ на основе n -ого количества принятых шифртекстов системой от других участников. Отсутствие зависимости от открытых текстов закономерно приводит к аналогичной теоретической доказуемости и указывает на возможность применения системой нескольких различных алгоритмов.

Пример 3.2. Отправление всех шифртекстов $C = \{c_1, c_2, \dots, c_n\}$ сети за один раз также является алгоритмом $C \leftarrow A(C)$, хоть и специфичным по своей природе, т.к. исключает какую бы то ни было интерактивность в лице запросов и ответов. Данный пример интересен и тем, что с некоторой вероятностью все шифртексты во множестве C обязательно должны оставаться ложными. В противном случае, будут нарушаться сразу два условия: неразличимости, где известным будет становиться неравенство $C^k \neq \emptyset$, и систематичности, где алгоритм генерации будет зависим от открытых текстов: $A(C) = A(m)$.

Пример 3.3. Небольшое различие алгоритма A при генерации истинных и ложных шифртекстов может нарушить систематичность генерации. Предположим, что задан статичный период t , как условие алгоритма, но при этом генерация шифртекстов c_i занимает продолжительное время x , такое что $0 < x < t$. Далее, если предположить, что ложные шифртексты генерируются на моменте отправления ($c_i = E_r(v_i) \leftarrow A(t)$), в обход очерёдности и по причине отсутствия в очереди каких-либо сообщений, тогда шифртекст c_i отправится спустя время равное $x+t$. В свою очередь, если истинные сообщения генерируются до момента отправления, т.е. сначала помещаются в очередь ($c_i = E_k(m_i) \leftarrow A(tQ)$), то время их отправления будет равно t на следующем этапе генерации. Таким образом, систематичность

алгоритма нарушается разными вводными условиями (независимыми событиями) при генерации истинных и ложных шифртекстов соответственно.

Пример 3.4. Различие между алгоритмами $A(t)$ и $A(tQ)$ определяется лишь отсутствием или существованием очереди сообщений. Теоретически оба алгоритма основаны на независимых событиях, но практически они отличаются тем, что при $A(t)$ генерация истинных шифртекстов будет более затруднительной, т.к. она приводит к ручному их созданию на конкретном периоде времени t по причине отсутствия механизма отложенности сообщений до заданного условия. Алгоритм с очередью $A(tQ)$, напротив, позволяет генерировать и сохранять сообщения в любой момент времени, вне зависимости от заданного периода t . Именно поэтому QV-задача именуется как Queue, а не Time Based.

Далее, в QV-сетях предполагается использование асимметричной криптографии по умолчанию, и как следствие ключи $k_i \neq k_i^{-1}$ не связаны между собой напрямую. Тем не менее QV-сети вполне способны руководствоваться исключительно симметричной криптографией при которой $k_i = k_i^{-1}$. В таком случае ключи будут представлять не каждого отдельного участника системы из n возможных, а непосредственно связь между её участниками из $n(n - 1)/2$ возможных рёбер графа (системы), что приводит также к появлению общих ключей вида $k_i = k_j$ для некоторых i, j участников. Если в системе заложен механизм маршрутизации, то для расшифрования получатель должен будет использовать уже не один конкретный ключ со стороны отправителя, а все ему известные ключи посредством метода их перебора.

Использование симметричной криптографии может повысить криптостойкость QV-net системы в условиях подготовки к постквантовой криптографии или в результате её наступления, т.к. известно, что современные симметричные алгоритмы с большой длиной ключа (256 бит и более) являются квантовоустойчивыми [5, с.131], в то время как новые разрабатываемые асимметричные алгоритмы, с консервативной точки зрения, ещё слишком новы, чтобы считаться безопасными, в том числе даже для классических компьютеров. Использование симметричной криптографии, в свою очередь, несёт также и ряд издержек:

1. Усложняется общая система количественного хранения и распространения ключей: $2n$ в асимметричной криптографии, $n(n - 1)/2$ в симметричной криптографии [3, с.278],
2. Расшифрование данных будет приводить к линейному перебору всех известных симметричных ключей k_i , что при наличии асимметричных ключей требовало лишь одного ключа k^{-1} ,
3. Понизится модель угроз с возможностями активного посредника подменять публичные ключи до возможности пассивного посредника просматривать секретные ключи [3, с.80],
4. Имитовставки, в отличие от цифровых подписей асимметричного раздела криптографии, не позволяют однозначным образом подтверждать авторство сообщений [6, с.46],

2.1. Недостатки QВ-сетей

К сожалению QВ-сети неидеальны и также обладают, свойственными своему классу, проблемами и недостатками, ряд из которых приводит к ограничению прикладного использования, другой ряд приводит к проблемам доступности сети:

1. Линейная нагрузка на сеть. В QВ-сетях каждый отправляет сообщение всем с той лишь целью, чтобы невозможно было сузить область реальной коммуникации участников системы. Алгоритмом маршрутизации становится слепая (заливочная) маршрутизация [7, с.398], вследствие чего увеличение количества узлов сказывается линейно $O(n)$ на увеличение нагрузки всей системы,

2. Привязанность к очередности. Каждый узел в QВ-сети так или иначе завязан на собственной очередности сообщений Q , где каждый период времени равный t зашифрованное сообщение отправляется в сеть. Это значит, что повысить пропускную способность узла возможно лишь в трёх сценариях, каждый из которых будет приводить к увеличению нагрузки на всю сеть:

1. Повысить размер передаваемых открытых сообщений m_1, m_2, \dots, m_n ,
2. Повысить количество отправляемых шифртекстов за раз $c_1, c_2, \dots, c_n \leftarrow^t Q$,
3. Понизить период генерации сообщений t ,

3. Связность абонентов коммуникации. QВ-сети не предполагают анонимности между узлами непосредственно участвующих в общении. Связано это в первую очередь с тем, что в QВ-сетях отсутствует такое понятие как полиморфизм информации [2, с.62], то есть состояние информации в системе при котором её внешний вид постоянно меняется от узла к узлу, как для внутренних, так и для внешних наблюдателей. Такое свойство позволяет разрывать связь между отправителем и получателем посредством передаваемого объекта, т.е. самой информации.

Вследствие всех вышеприведённых недостатков область применения QВ-сетей становится более ограниченной:

1. Из-за линейной нагрузки на сеть и привязанности к очередности QВ-сети плохо масштабируются и могут работать лишь в малых группах до N участников. Предел количества участников ограничен пропускной способностью самой сети, а также мощностью узлов постоянно шифрующих и расшифровывающих исходящий / входящий трафик. Вследствие этого недостатка реализация стриминговых сервисов и видео / аудио звонков становится либо очень затруднительной задачей, либо вовсе невыполнимой,

2. Из-за связности абонентов коммуникации ограничивается ряд прикладных решений в которых важна анонимность узлов друг к другу. Вследствие этого, появляется наиболее релевантная композиция QВ-сетей с F2F-сетями (friend-to-friend), где установление коммуницирующей связи происходит двумя абонентами системы, а не одним из. Это не решает проблему отсутствия анонимности между связываемыми узлами, но даёт дополнительную защиту от несогласованного автоматического связывания и более явную связь доверительных коммуникаций, предполагающую что ни один из абонентов не будет пытаться деанонимизировать другого.

Алгоритм 2. Фильтрация сообщений в F2F-сетях на языке псевдокода

ВВОД: множество друзей F , отправитель s , функция-обработчик h , сообщение m

ВЫВОД: обработанное сообщение $h(m)$ ИЛИ завершение алгоритма

```
if ( $s \notin F$ ) {  
    return (* Завершение алгоритма *)  
}  
return  $h(m)$  (* Обработка сообщения *)
```

2.2. Активные наблюдения

Из всего вышеописанного ранее было доказано, что QV-задача невосприимчива к любым пассивным наблюдениям. Это говорит о том, что данная задача принадлежит классу задач анонимизации с теоретической доказуемостью. В свою очередь, теоретическая доказуемость не является абсолютной, т.к. сводится лишь к решению проблем пассивных наблюдений, игнорируя и обходя активные. Как будет показано далее, QV-задача не обладает абсолютной анонимностью, и может быть уязвима к конкретным активным наблюдениям.

В QV-сетях отсутствует свойство полиморфизма информации, вследствие чего становится возможным ряд активных наблюдений, когда один из абонентов коммуникации желает скрыть выстроенную им связь от собеседника. Полиморфизм в анонимных сетях чаще всего достигается множественным шифрованием, где при передаче от одного узла к другому постепенно снимаются наложенные слои шифрования, как например в Tor, I2P, Mixminion [8][9][10]. Такое свойство позволяет разграничивать связь абонентов друг к другу, тем самым, анонимизируя их.

$$E_{k3}(E_{k2}(E_{k1}(m))) \rightarrow E_{k2}(E_{k1}(m)) \rightarrow E_{k1}(m) \rightarrow m$$

Так например, если предположить, что в QV-сети будет существовать роль в виде ретрансляторов, скрывающая сетевые адреса абонентов (IP-адреса) друг от друга посредством перенаправления трафика, и будет при этом существовать кооперация одного из абонентов с глобальным наблюдателем, то задача связывания $IP \leftrightarrow k_i$ будет тривиальной, т.к. абоненту достаточно будет получить одно истинное сообщение $m = D_k^{-1}(c)$ от собеседника, а далее по полученному шифртексту $c = E_k(m)$ глобальный наблюдатель сможет определить первое его появление, тем самым деанонимизировав отправителя или получателя.

Ситуацию можно усложнить для наблюдателей при помощи добавления канального шифрования, как например в Crowds [11]. В таком случае глобальный наблюдатель не сможет явно связать отправленное и полученное сообщение, потому как оно будет постоянно менять свой вид при передаче от одного узла к другому:

$$E_{k3}(m) \rightarrow E_{k2}(m) \rightarrow E_{k1}(m) \rightarrow m$$

Тем не менее это не является полиморфизмом информации, т.к. не выполняет функцию разграничения узлов между собой к маршрутизирующей информации m . Вследствие этого, задачей глобального наблюдателя станет вживание подконтрольных узлов в систему рядом с каждым другим узлом. При таком сценарии он также легко сможет решить задачу $IP \leftrightarrow k_i$.

Чисто технически в QV-сеть можно внедрить и множественное шифрование, чтобы разграничивать абонентов друг от друга, но в таком случае:

1. Уменьшится скорость передачи информации, т.к. каждый маршрутизирующий узел должен будет сохранять полученное им ранее сообщение в свою очередь,
2. Усложнится система анонимизации в целом, т.к. вместо одной задачи анонимизации = QV будет использоваться уже композиция задач = QV + Onion,
3. Композиция задач QV + Onion обладает рядом тонкостей с более сложными активными наблюдениями [2, с.159], но всё также деанонимизирующими абонентов сети.

Теорема 4. При наличии активного внутреннего наблюдателя f_i в роли собеседника для узла i , задача анонимизации всегда будет сводиться к сокрытию связи между данными абонентами коммуникации.

Доказательство 4. Если пойти от обратного и предположить, что задача анонимизации может быть определена фактом скрывания коммуникации при наличии активного внутреннего наблюдателя f_i в роли собеседника для узла i , тогда мы придём к противоречию, т.к. атакующий при отправлении или получении сообщений от собеседника априори будет знать информацию о том, что в конкретные промежутки времени t_1, t_2, \dots, t_n , присутствовал обмен истинными текстами, а потому был нарушен сам факт скрывания коммуникации собеседника i . В результате, задача анонимизации начинает сводиться к поиску конкретной связи, а не к наличию факта существования этой связи.

Далее, если предположить сценарий атаки на QV-сеть при котором в кругу друзей $F = \{f_1, f_2, \dots, f_n\}$ участника i будет существовать злоумышленник f_j в роли активного наблюдателя способного отправлять запросы и получать ответы от i , тогда модель атаки будет сводиться к анализу состояния очереди Q_i . Предположим далее, что участник i выставил статичный период генерации сообщений равный t_i . В таком случае f_j сможет в определённые интервалы времени $\{t'_i, 2t'_i, \dots, nt'_i\}$, зависящие от периода времени $t_i \Rightarrow (kt'_i = kt_i + x)$, где $x \in [0; t_i)$, отправлять запрос $R_{kt'i}$ участнику i с целью анализа времени ответа. Если ответ, полученный после запроса $R_{kt'i}$, будет генерироваться в диапазоне dt_i , где $d > 1$, то это будет означать факт реальной коммуникации участника i с кем либо в сети в множестве периодов $D = \{(1+k)t_i, (2+k)t_i, \dots, (d-1+k)t_i\}$, т.к. для ответа потребовался более чем один период. Если же $d = 1$, тогда участник i ни с кем не кооперировал в период $(1+k)t_i$.

Таким образом, вышеописанная атака снижает качество анонимности QV-сетей с сокрытия факта активности до сокрытия коммуникационной связи между абонентами. Иными словами, при таком активном наблюдении теперь становится возможным определение состояния субъекта в лице отправления или получения истинных сообщений, но до сих пор остаются под вопросом следующие моменты:

1. С какими узлами общался прослушиваемый участник i в множестве периодов D ?
2. Являлся ли прослушиваемый участник i инициатором запросов при множестве D ?
3. Может ли участник i намеренно генерировать ложные сообщения в роли истинных?

Далее, если допустить ситуацию при которой у каждого узла $i \in \{1, 2, \dots, n\}$, не относящегося к наблюдателям, в друзьях будет присутствовать как минимум один активный наблюдатель $f_i \in \{f_1, f_2, \dots, f_n\}$, тогда первая задача может быть решена тривиальным образом, при условии, что атакующие будут находиться в кооперации, т.е. $\forall i, j, f_i = f_j$, а участники будут использовать тип связи «запрос-ответ». В таком случае близкая к одновременной загруженность очереди $(d+x)t_i$, где $d > 1, x \in \{0, 1, 2\}$ у нескольких участников сети i будет свидетельствовать об ограничении первоначального множества наблюдения за счёт разделения узлов i, j по заполненным $Q_i \neq \emptyset$ и пустым $Q_j = \emptyset$ очередям.

Решение второй задачи может базироваться на условиях первой, когда будет существовать множество активных и кооперирующих наблюдателей $f_i \in \{f_1, f_2, \dots, f_n\}$, а участники будут использовать тип связи «запрос-ответ». В таком случае, для поиска ответа наблюдателям потребуется выявлять факт начала коммуникации у всех узлов $i \in \{1, 2, \dots, n\}$ посредством перехода состояния очереди из пустого в заполненное: $Q_i = \emptyset \rightarrow Q_i \neq \emptyset$. Первый узел, осуществивший такой переход, с большей вероятностью становится инициатором запроса.

Решение третьей задачи является наиболее трудоёмким с точки зрения наблюдателей, т.к. во-первых, оно связано с решением задачи неразличимости шифртекстов $P_{(i)}$, что является вычислительно трудной задачей, во-вторых, отсутствие решения третьей задачи усложняет выявление закономерностей первых двух задач, посредством скрытия точного состояния очередей у абонентов коммуникации « $Q_i = \emptyset?$ ». Хотя такая задача и связывает себя с задачей неразличимости, она ей не тождественна, т.к. помимо неё присутствует и задача систематичности $P_{(s)}$. Вследствие этого, чтобы доказать безопасность третьей задачи, необходимо свести её сложность к задаче деанонимизации $P_{(d)}$. Но это невозможно, т.к. ответ наблюдателю становится процедурой нарушения независимости события x в задаче систематичности. За счёт этого, наблюдатели могут быть уверены, что в момент ответа dt_i участник i не мог параллельно отвечать другому абоненту. Таким образом, третья задача не способна дать гарантий анонимности и посредством более длительного наблюдения всё также может проявлять паттерны, свойственные решению первой задачи.

Проблему можно решить расширением количества очередей $Q \rightarrow \{Q_1, Q_2, \dots, Q_n\}$ на одном узле с их привязкой к абонентам коммуникации $i \in \{1, 2, \dots, m\}$. При этом, количество очередей должно быть априори задано статичным значением n , чтобы невозможно было выявить количество связанных друзей m , т.е., количество друзей m всегда должно иметь сравнение: $m \leq n$. В результате, расширение очередей приведёт либо к повышению количества отправляемых шифртекстов за раз: $c_1, c_2, \dots, c_n \leftarrow^t Q_1, Q_2, \dots, Q_n$, либо к повышению периода генерации шифртекстов: $c_i \leftarrow^{nt} Q_i$. Это, в свой черёд, будет являться гарантией / доказательством того, что активный наблюдатель не сможет повлиять на очередь сообщений других абонентов, т.к. все его действия будут привязаны и ограничены одной непересекающейся очередью.

2.3. Сравнение с другими задачами

По своим характеристикам QB-задача наиболее близка к DC-задаче из-за следующих особенностей: теоретически доказуемая анонимность, периодичность генерации сообщений, принадлежность к второму вектору развития анонимных коммуникаций [2, с.71], сложность масштабирования. Отличия QB от DC-сетей, с положительной точки зрения, присутствуют

следующие: периодичность генерации может иметь динамичную величину, анонимность не зависит от выстроенных связей с другими участниками, более простая программная реализация. Негативное отличие определяется отсутствием полиморфизма информации. Более детальное и общее сравнение QV-задачи с другими задачами анонимизации как теоретическими, так и практическими представлено в *Таблице 1*.

	QV	EI	DC	Onion	Proxy
Теоретическая доказуемость	+	+	+	-	-
Накопительный эффект анонимности	-	+	-	-	-
Полиморфизм информации	-	+	+	+	-
Вероятностная маршрутизация	-	+	-	+/-	+/-
Периодичность генерации сообщений	+/-	-	+	+/-	+/-
Независимость анонимности от связей	+	-	-	-	-
Простота масштабирования	-	-	-	+	+
Простота программной реализации¹	+	-	-	+/-	+
Стадия анонимности	5 [^]	6	1 [^]	4 или 6	3
Сеть-представитель	Hidden Lake	-	Herbivore	Tor	Crowds

Таблица 1. Сравнение задач анонимизации

3. Функция шифрования

Как было ранее показано, QV-сети зависимы от качества функции E и ключей k, r шифрования. Качество ключей шифрования определяется в первую очередь качеством ГСЧ (генератором случайных чисел) и/или КСГПСЧ (криптографически стойким генератором псевдослучайных чисел). Анализ таковых генераторов сложен по причине разных сред, в которых они исполняются, и средств, которые они задействуют в ходе своего выполнения [6, с.190]. Поэтому исходя из логики абстрагирования мы будем далее предполагать, что ключи генерируются качественным и безопасным образом, фокусируясь тем самым исключительно на логике исполнения функции шифрования.

$$E_{(k, \text{privA}, \text{pubB})}(m) = E''_k(E'_{(\text{privA}, \text{pubB})}(m))$$

Функция шифрования в сети Hidden Lake состоит из двух этапов, каждый из которых выполняет свою точно заданную роль. Первый этап $E'_{(\text{privA}, \text{pubB})}$ сводится к непосредственному и первичному шифрованию данных с целью их сокрытия от посторонних лиц, используя для этого гибридную схему шифрования (асимметричная + симметричная

¹ Характеристика сложности может быть определена количественным соотношением суммы успешных пассивных / активных наблюдений к сумме необходимых процедур для предотвращения от таковых наблюдений. Если сумма процедур определяется небольшим числом, тогда простота программной реализации будет минимальна, что, тем не менее, не говорит об итоговой безопасности сети, т.к. малое количество процедур может свидетельствовать и об отсутствии принимаемых мер при ликвидации успешных наблюдений. В таком случае, простота программной реализации может быть не только следствием высокой безопасности, но и следствием пониженной модели угроз.

криптография) [2, с.125]. Вторым этапом E''_k сводится к разделению нескольких сетей посредством применения разных ключей шифрования (сетевых ключей).

3.1. Первый этап шифрования

$$E'_{(privA, pubB)}(m) = (E_{pubB}(k') \parallel E_k(pubA \parallel s \parallel h \parallel S_{privA}(h) \parallel m')),$$

$$h = H_{mac(s)}(pubA \parallel pubB \parallel m'), m' = f(m), k' = [RNG], s = [RNG],$$

где k' - сеансовый ключ шифрования рассчитанный на одно сообщение, s - криптографическая соль рассчитанная на одно сообщение, m - открытое сообщение, $pubA$, $pubB$ - публичные ключи участников A , B соответственно, $privA$ - приватный ключ участника A , h - результат хеширования, S - функция подписания, f - функция дополнения сообщения до константной величины. В этой схеме предполагается, что A - есть отправитель информации m , B - есть получатель данной информации.

Безопасность данной функции зависит непосредственно от публичного ключа шифрования $pubB$, которым шифруется последующий сеансовый ключ k' , от качества ГСЧ / КСГПСЧ которым был сгенерирован k' , и также от безопасности самих функций шифрования E_{pubB} , E_k .

Данная схема интересна тем, что она скрывает всю информацию в зашифрованной оболочке, не позволяющей осуществлять атаки на идентификацию отправителя или получателя. Так например, если бы хеш-значение h и подпись $S_{privA}(h)$ не находились в зашифрованном блоке E_k , тогда была бы возможна атака анализа зашифрованных сообщений по уже имеющемуся списку публичных ключей $\{pub_1, pub_2, \dots, pub_n\}$, проверяющих их аутентичность $V_{pubi}(S_{privA}(h)) = h$.

Далее, если была бы известна криптографическая соль s и хеш-значение h , то можно было бы составить таблицу наиболее часто встречающихся сообщений $\{m_1, m_2, \dots, m_n\}$ с различными комбинациями участников i, j из множества всех узлов сети N по равенству $H_{mac(s)}(pub_i, pub_j, f(m_i)) = h$.

Плюс к этому, данная схема является самодостаточной [2, с.121] на сетевом уровне работы QV-сетей в контексте заливочной маршрутизации, потому как позволяет обеспечивать идентификацию субъектов лишь и только при помощи асимметричной криптографии. Определить отправителя сообщения становится возможным посредством корректного расшифрования, т.е. при условии, когда получатель зашифрованного сообщения располагает нужным приватным ключом.

Данный этап также предполагает, что сообщение $f(m)$ имеет статичную величину. Иными словами, при каждом вызове функции шифрования $E'_{(privA, pubB)}$, для всех m_i, m_j из $\{m_1, m_2, \dots, m_n\}$ соблюдается длина сообщения L от функции l , такая что $l \Rightarrow l(f(m_i)) = l(f(m_j)) = L$. Это становится возможным за счёт процедуры препроцессинга f , ограничивающей длину входного сообщения величиной L , и дополняющей длину входного сообщения до L . Целью такой процедуры становится защита от атак по анализу размера сообщений, при которых может выявляться структура передаваемого сообщения. Например, запросы чаще всего по размеру меньше чем ответы, передаваемые видео или аудио -файлы часто по размеру больше, чем обычные текстовые сообщения, системные / автоматические запросы меньше по размеру, чем вручную выполненные и т.д [12].

3.2. Второй этап шифрования

$$E''_k(m) = E_k(p(h) || h || mv),$$

$$h = H_{mac(k)}(mv), mv = (m || v), v = [RNG],$$

где k - ключ сети, p - функция доказательства работы, h - результат хеширования, m - открытое сообщение, v - пустые байты случайной длины. Функция придерживается MtE подхода (MAC then Encrypt), где сначала вычисляется MAC (Message Authentication Code), а далее сообщение mv , полученный код h и доказательство $p(h)$ шифруются функцией E_k .

Данный этап шифрования выполняет несколько задач:

1. Разграничивает разные сети по ключу сети k , чтобы их нельзя было слить в одну общую систему. Это достигается преимущественно за счёт функции доказательства работы p , т.к. из-за неё становится более затратным перешифровывать весь трафик направленный из одной сети с ключом k_1 в другую сеть с ключом k_2 ,

2. В отличие от первого этапа шифрования, придающего информации m статичный вид при помощи функции f , второй этап напротив - придаёт информации случайный размер v с целью противодействия блокировкам, направленных на анализ размерности сообщений и их структуру.

Доказательство работы p определяется алгоритмом proof-of-work (PoW) [13], где для конкретного хеш-значения h необходимо найти такое число i , чтобы результат $h_i = H(h || i)$ представлял собой битовый вектор с определённо заданным n -ым количеством нулей в качестве префикса, пример $00000000(n)...11001010$. Число n именуется сложностью работы.

Стоит также заметить, что ключ шифрования k может являться открытым параметром, если отсутствует необходимость противодействовать блокировкам. В таком случае, вторая задача становится побочной (опциональной), а ключ k - просто общеизвестной настройкой для разграничения сетей.

Подход Encrypt-then-MAC (EtM) не применяется в схеме шифрования второго этапа по двум причинам:

1. Используется один ключ k для шифрования и аутентификации вместо двух ключей k_1, k_2 для этих задач. Если применить в этой ситуации подход EtM, тогда на один и тот же ключ k будет открыто два вектора нападения, вместо одного. Эту проблему можно было бы решить при помощи использования KDF (функции формирования ключей), которая бы позволила из одного ключа создать несколько. Тем не менее это усложнит общую схему шифрования, а также откроет дополнительный вектор нападения на саму KDF,

2. Учитывается принцип Хортонa: «аутентифицировать нужно не то, что сказано, а то, что имеется в виду» [6, с.130]. При успешной атаке на функцию вычисления MAC в подходе EtM, либо при неправильном распределении ключей k_1, k_2 может возникнуть ситуация когда аутентификация будет выдавать положительный результат на зашифрованное сообщение, но сама процедура расшифрования будет некорректной. Если сообщение представляет собой хаотичный набор битов, то мы никогда не узнаем его истинность.

У подхода MtE безусловно существует недостаток в том, что перед тем как проверить целостность и аутентичность информации необходимо её расшифровать. У подхода EtM такой проблемы нет. Вследствие этого был также сформирован принцип криптографической обреченности Марлинспайком, который гласит: «если вы вынуждены выполнить любую криптографическую операцию до проверки имитовставки полученного сообщения, то это так или иначе, но неизбежно приведет к роковому концу» [14, с.93]. Данный принцип противопоставляется принципу Хортона, когда речь заходит о выборе одного из двух подходов: MtE или EtM. Тем не менее выбор EtM обусловлен ещё тем, что принцип криптографической обреченности нарушается также на моменте первого, и куда более затратного, этапа шифрования.

4. Сетевое взаимодействие

Наличие QB-задачи предполагает, что каждое отправляемое сообщение в сети будет достигать всех её участников. При этом в такой задаче ничего не говорится о том, как сообщение будет достигать узлов при отсутствии связи «все-ко-всем», какой протокол передачи данных будет использоваться, как должны работать прикладные приложения в такой модели, как децентрализованная структура будет обходить NAT в сети Интернет и т.д. Все эти вопросы требуют дополнительного пояснения исходя из конкретной реализации сетевого взаимодействия.

4.1. Микросервисная архитектура

Анонимная сеть HIDDEN Lake, как приложение, представляет собой набор сервисов, каждый из которых выполняет свою конкретную задачу [15]. Выбор микросервисной архитектуры, в противовес монолитной, был сделан с учётом следующих аспектов:

1. Микросервисная архитектура позволяет упростить и децентрализовать разработку прикладных сервисов, благодаря чему становится возможным применение различных языков программирования и технологий при реализации собственных приложений,
2. Микросервисная архитектура позволяет разделять ответственность сервисов к обрабатываемой или хранимой информации, благодаря чему при усложнении системы, корреляция самого усложнения будет минимально распространяться на сервисы,
3. Микросервисная архитектура позволяет добавлять новые функции, редактировать их или удалять вовсе без необходимости перекомпиляции и перезагрузки всех сервисов, что может положительно сказаться как на тестировании, так и на отказоустойчивости.

Ядром сети Hidden Lake является сервис HLS (service), который непосредственно исполняет QB-задачу и все функции шифрования / расшифрования соответственно. Помимо сервиса HLS, в сети Hidden Lake также существует ряд прикладных сервисов, по типу HLM (messenger), HLF (filesharer), HLR (remoter), ряд сервисов-помощников, по типу HLT (traffic), HLL (loader), HLE (encryptor), а также ряд адаптеров HLA, привязанных на текущий момент к сервисам common (тестовый сервис) и chatingar. В результате этого, сеть Hidden Lake можно представить как композицию нескольких сервисов.

$$HLS = D \times QB-net [E_{(k,privA,pubB)}(m) = E''_k(E'_{(privA,pubB)}(m))]$$

$$Hidden-Lake = \Sigma_{i=1}^n APP_i \times HLS \times (HLT \times \Sigma_{j=1}^m HLA_j)^t$$

где APP - множество прикладных сервисов, HLA - множество адаптеров к сторонним сервисам, D – доставщик открытых сообщений к конкретному прикладному сервису, $t = 0$ or 1 - параметр определяющий отсутствие / существование ретрансляции сообщений. При отсутствии прикладных приложений, факта ретрансляции, и как следствие, адаптеров, сеть Hidden Lake становится равной сервису HLS: $Hidden-Lake = HLS$. Это есть минимальная характеристика, при которой HL ещё остаётся собой. При удалении же сервиса HLS, система перестаёт являться Hidden Lake сетью, т.к. лишается своего ядра.

Вследствие всего вышеописанного реализация сети Hidden Lake может представлять разные режимы коммуникации: классический, ретрансляционный и адаптационный.

1. Под классическим режимом коммуникации понимаются прямые или косвенные соединения между HLS узлами без использования сторонних сервисов-помощников. В такой модели каждый участник является и производителем шифртекстов, и их ретранслятором,

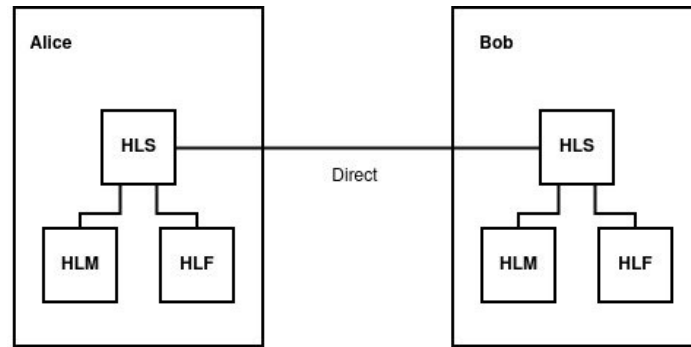


Рисунок 2. Классический режим. $Hidden-Lake = (HLM+HLF) \times HLS$

2. Под ретрансляционным режимом коммуникации понимаются соединения между HLS узлами с применением ретрансляционных узлов HLT. Такой режим коммуникации позволяет обходить NAT посредством использования сервисов HLT в роли выделенных TURN-серверов. При этом централизованный характер серверов никак не будет влиять на качество анонимности из-за существования теоретической доказуемости в лице QB-задачи,

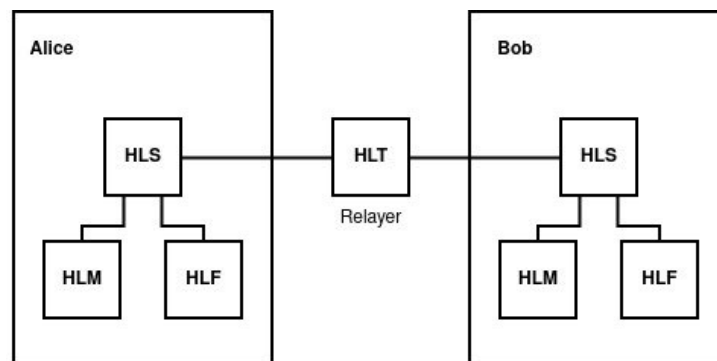


Рисунок 3. Ретрансляционный режим. $Hidden-Lake = (HLM+HLF) \times HLS \times HLT$

3. Под адаптационным понимаются соединения между HLS узлами с применением HLA сервисов. Такой режим коммуникации позволяет внедрять анонимизированный трафик в инородные / внешние системы, не связанные как-либо с сетью Hidden Lake. Для подобного режима используется также сервис-помощник HLT, позволяющий транслировать TCP трафик, получаемый на выходе от HLS, в HTTP трафик, отправляемый на вход HLA.

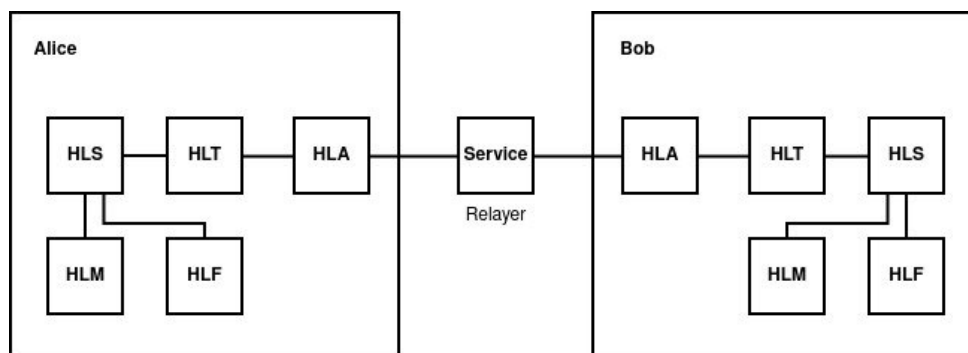


Рисунок 4. Адаптационный режим. Hidden-Lake = (HLM+HLF) × HLS × HLT × HLA=Service

4.2. Стек протоколов «GP/12»

Абстрактный характер анонимной сети Hidden Lake формируется стеком протоколов GP/12 (сокращение от go-peer² и 1,2 – этапы шифрования), аналогичным по своей сути стеку протоколов TCP/IP. В нём также присутствуют четыре уровня: канальный (CL), сетевой (NL), транспортный (TL) и прикладной (AL), но при этом имеется ряд следующих отличий:

1. GP/12 идентифицирует узлы по публичным ключам, а не по IP адресам,
2. GP/12 может использовать задачу теоретически доказуемой анонимности,
3. GP/12 не зависит от сетевых протоколов и систем коммуникаций.

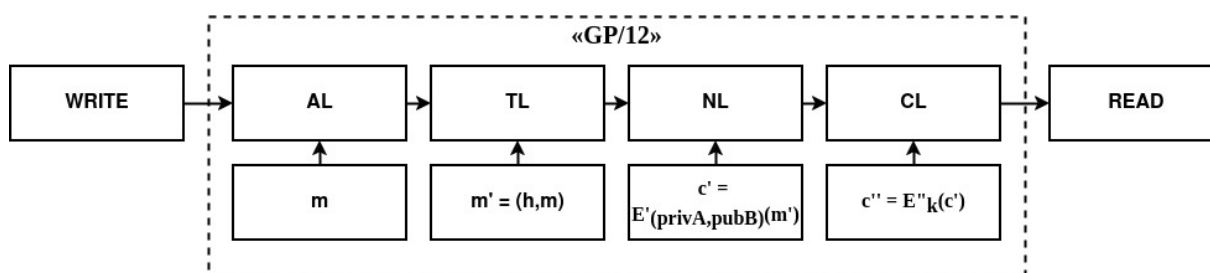


Рисунок 5. Стек протоколов GP/12

Канальный уровень в сетевой модели GP/12 представляет собой распространение сообщений методом заливочной маршрутизации. Данный уровень характеризуется применением второго этапа шифрования, когда важен сам факт успешной передачи

² Исходный код анонимной сети Hidden Lake содержится в репозитории проекта go-peer: https://github.com/number571/go-peer/tree/master/cmd/hidden_lake. Для сети Hidden Lake проект go-peer является фреймворком по причине их общей принадлежности к кодовой базе и релизным версиям. В go-peer содержатся все основные модули: функции шифрования, работа с очередью сообщений, сетевое взаимодействие. Hidden Lake, в свою очередь, активно применяет данные компоненты для реализации таких сервисов, как HLS, HLT, HLE.

сообщения всем своим соединениям. Сетевой уровень представляет собой распространение сообщений по конкретным узлам сети, используя для этого публичные ключи в роли идентификаторов. Данный уровень характеризуется применением первого этапа шифрования, когда важна конфиденциальность и аутентичность сообщений. Транспортный уровень представляет собой маршрутизацию открытых сообщений (успешно расшифрованных) по конкретно заданным прикладным сервисам. Прикладной уровень представляет собой принятие открытого сообщения с дальнейшей его обработкой. Таким образом, стек протоколов GP/12 может быть изображён, как композиция четырёх уровней: $AL \times TL \times NL \times CL$.

Вследствие вышеописанного, исполнение полного стека протоколов GP/12 от канального до прикладного уровней возможно лишь с использованием минимум двух сервисов в сети Hidden Lake, т.к. HLS, являясь ядром сети, покрывает лишь первые три уровня: CL, NL, TL, в то время как последний уровень AL может быть покрыт лишь использованием прикладных приложений по типу: HLM, HLF, HLR.

Также, сетевая модель GP/12 необязательно должна обладать свойством анонимности из-за чего сервис HLS, при удалённой QB-задаче с целью сохранения GP/12 совместимости, может быть заменён композицией из трёх сервисов: $HLT \times HLE \times D$, предоставляющих первый, второй и третий уровни соответственно. Именно таким образом было создано приложение secpy-chat³.

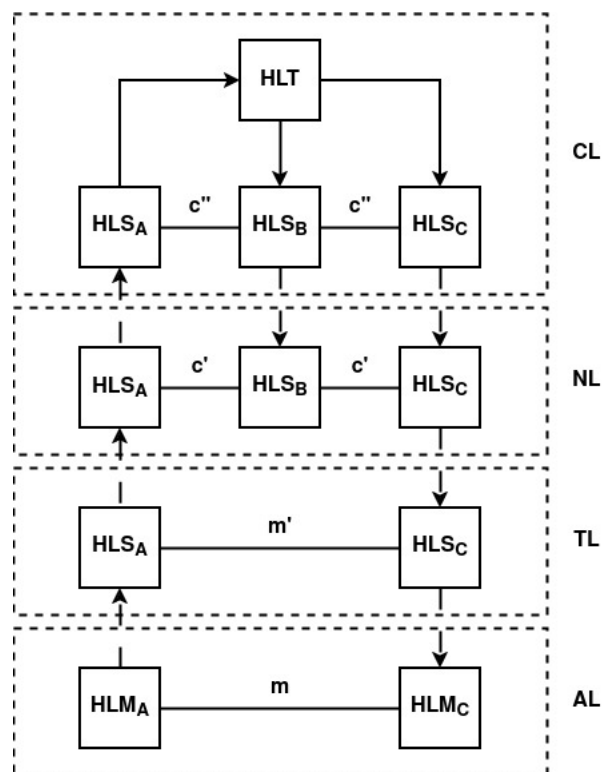


Рисунок 6. Модель GP/12 на примере сети Hidden Lake в режиме ретрансляции трёх узлов, где А – отправитель, С – получатель

Анонимизация трафика может быть сформирована на втором уровне стека GP/12, когда генерация сообщений будет привязана к какой-либо задаче. Это может быть как QB, так и

³ Исходный код клиент-безопасного приложения secpy-chat содержится в репозитории проекта go-peer: https://github.com/number571/go-peer/tree/master/cmd/secpy_chat

DC, EI –задачи. Основное ограничение в таком выборе заключается в необходимости существования теоретической доказуемости, без которой невозможно будет далее формировать абстрактную систему со свойством анонимности. Если будут применяться Proxu или Onion –задачи, то в таком случае сеть на базе сетевой модели GP/12 не станет более анонимной, потому как этим задачам, для своего корректного исполнения, требуется немонолитная система, что противоречит определению абстрактных анонимных сетей, которым неважна системная централизация.

Для своей реализации стеку протоколов GP/12 достаточно лишь существования сетевых интерфейсов WRITE и READ, вследствие чего он может адаптироваться далее к конкретной среде коммуникации. Так например, модель GP/12 в сети Hidden Lake по умолчанию базируется на протоколе транспортного уровня TCP, а при помощи сервисов HLA (адаптеров) может базироваться как поверх протоколов прикладного уровня: HTTP, SSH, FTP и т.п. модели TCP/IP [16], так и вовсе без TCP/IP, как пример, используя радиовещание или систему видимого света.

5. Структурные параметры

Математические модели позволяют выявить корректность общей логики работы, но не позволяют определить безопасность конкретной реализации. Так например, мы можем лишь предполагать, что какая-либо функция или принимаемое ей значение будет безопасным. Тем не менее всё это ничего не говорит о конкретной реализации, выбираемых параметрах, модели угроз и подходах проектирования. Таким образом, необходимо уделить внимание не только общему описанию работы сети, но и подробному изложению её структурных параметров.

1. Реализация сети Hidden Lake придерживается принципа наименьшего количества зависимостей: «Если какую-то функцию можно реализовать относительно легко стандартной библиотекой языка, то следует отдать предпочтение именно такому подходу, даже если итоговое решение не будет являться лучшим из множества возможных». Данный принцип позволяет уменьшить количество используемых зависимостей, тем самым сужая вектор атак на сеть через сторонние реализации библиотек и приложений,

2. Симметричная функция шифрования E_k определяется блочным алгоритмом AES с длиной ключа 256 бит в режиме шифрования CFB (режим обратной связи по шифртексту), где $c_0 = IV$, $c_i = E_k(c_{i-1}) \text{ xor } m_i$.

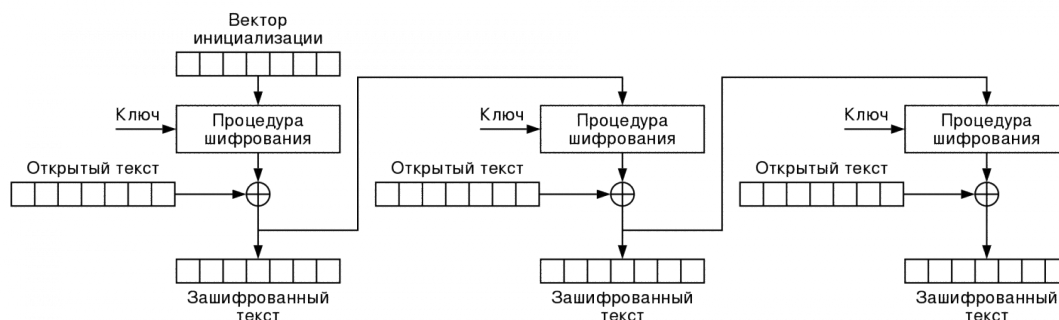


Рисунок 7. Режим шифрования CFB

Данный режим шифрования был выбран с учётом следующих моментов:

1. Режиму шифрования CFB не требуются дополнения (padding), как это требуется например режиму шифрования CBC (режим сцепления блоков шифртекстов). Вследствие этого упрощается выставление статичного размера шифрованного сообщения, а также ликвидируются возможные атаки оракула [17],
 2. Режим шифрования CFB не является поточным режимом шифрования, как например OFB (режим обратной связи по выходу) или CTR (режим счётчика). Вследствие этого, у CFB отсутствует проблема в лице повторяемости гаммы. Если IV (вектор инициализации) повторится, то это приведёт к куда меньшим проблемам безопасности, чем при OFB, CTR режимах [6, с.93],
 3. Режим шифрования CFB не требует от алгоритмов шифрования соблюдения точных параметров, как например, 128-битные блоки при режиме GCM (режим счётчика с аутентификацией Галуа) [18, с.190], а также не требует от алгоритмов функции расшифрования, что позволяет не только упростить заменяемость небезопасных шифров, но и использовать в качестве процедуры шифрования односторонние функции [19, с.282],
 4. Не использовался режим шифрования GCM также и по причине излишних операций аутентификации и хранения токенов. Первый и второй этапы шифрования используют разные способы аутентификации сообщений. Вследствие этого, в плане гибкости использования, режим CFB становится более предпочтительным,
3. Асимметричная функция шифрования E_{pub} определяется алгоритмом RSA с длиной ключа 4096 бит со схемой кодирования OAEP (оптимальное асимметричное шифрование с дополнением). Асимметричная функция подписания S_{priv} определяется также алгоритмом RSA-4096, но уже со схемой кодирования PSS (схема вероятностной подписи). Для асимметричных функций расшифрования D_{pub} и подписания S_{priv} используется один ключ. По этой причине для одного ключа используются разные схемы кодирования [6, с.257],
4. При сравнении криптографических алгоритмов и схем на базе эллиптических кривых с алгоритмом RSA, может возникнуть сомнение о выборе последнего в качестве основного инструмента процедур идентификации, аутентификации и конфиденциальности. Для сохранения того же уровня безопасности можно было бы использовать ключи меньших размером, тем самым, уменьшив размер заголовков и увеличив максимальный размер допустимой полезной нагрузки в статичном сообщении. В конечном итоге, выбор RSA стал следствием следующего списка условий:
1. Сеть Hidden Lake написана на языке программирования Go [20], в стандартной библиотеке которого по умолчанию уже присутствует реализация алгоритма RSA со схемами кодирования OAEP, PSS. С точки зрения безопасности и принципа наименьшего количества зависимостей выбор такого алгоритма становится наиболее предпочтительным,

2. В сообществе языка Go отсутствует общепринятый на практике (как в стандартной, так и в сторонней библиотеках) и проверенный на безопасность пакет со стандартизированной схемой ECIES (интегрированная схема шифрования на эллиптической кривой),

3. В новых версиях Go, начиная с go1.20, появился протокол Диффи-Хеллмана на эллиптических кривых (ECDH), который смог бы лечь в основу кастомной реализации ECIES. Но, в таком случае, собственная реализация ECIES потребовала бы осуществление дополнительных действий и усилий, нацеленных на проверку корректности и безопасности написанного кода,

5. Первый этап шифрования не представляет собой попытку какого-либо скрывания структуры сообщения. Такая задача возлагается лишь на второй этап шифрования в целях разделения логики на сохранение статичного и приобретение динамичного размера в пределах одного сообщения,

6. Предполагается, что ключ сети на втором этапе шифрования редко меняется, обладает высокой энтропией и не является паролем. Следовательно, ключ сети не проходит сквозь какую бы то ни было функцию формирования ключа (KDF), но пропускается сквозь алгоритм хеширования SHA-256 для сжатия до фиксированного размера в 32 байт, чего требует алгоритм шифрования AES-256,

7. Ключи размером в 256 для AES и в 4096 бит для RSA соответственно были выбраны с консервативной точки зрения. Ключ в 256 бит AES сможет успешно противостоять постквантовой криптографии [5, с.131]. Ключ в 4096 бит RSA сможет противостоять постквантовой криптографии лишь на начальных этапах, потому как потребует от квантового компьютера примерно $2n+3 = 8195$ хорошо связанных между собой кубитов [5, с.134]. Тем не менее размер в 4096 бит RSA был выбран также при сравнительном анализе безопасности с длиной ключа симметричного алгоритма к атакам полного перебора, где $4096\text{bit RSA} \approx 140\text{bit}$ [21, с.54]. На текущий момент времени минимальной криптостойкостью считается длина ключа в 112 бит для симметричного алгоритма,

Symmetric Key Size (bits)	RSA and Diffie-Hellman Key Size (bits)	Elliptic Curve Key Size (bits)
80	1024	160
112	2048	224
128	3072	256
192	7680	384
256	15360	521

Таблица 2. NIST: сравнение длин ключей при лобовой атаке

8. В качестве алгоритма имитовставки (MAC) используется HMAC-SHA-256. Безопасность HMAC зависит от используемой им хеш-функции [18, с.168]. Безопасность SHA-256 может быть определена стойкостью к коллизиям в 128 бит, исходя из атаки парадокса дней рождения [6, с.52],

9. Ни первый этап шифрования, ни второй не защищают от атаки воспроизведения повторных сообщений [6, с.279], где спустя определённый период времени t злоумышленником может быть транслировано вновь, сохранённое ранее им же, шифрованное сообщение. Такое сообщение будет полностью верным, т.к. оно было сгенерировано истинным отправителем. Для защиты от подобного вида атак, сеть Hidden Lake поступает наиболее простым и радикальным способом, сохраняя хеши сообщений в свою локальную БД. Такой подход позволяет полностью исключить возможность успешного принятия дубликатов, в том числе и при перезагрузке узла, а также упростить в общем процедуру дедупликации, посредством исключения дополнительных проверок сообщений в пределах временных окон.

Подход сохранения хешей всех ранее принятых сообщений также имеет и свой минус в лице стремления к постоянному увеличению объёма базы данных. Как только появляется БД хешей, её ни в коем случае нельзя удалять или очищать, иначе все ранее принятые сообщения вновь смогут повториться, если злоумышленник будет применять атаку воспроизведения повторных сообщений. В такой парадигме вопрос начинает базироваться на объёме и частоте постоянно генерируемой информации. Хеш сообщения в сети Hidden Lake - это хеш-значение $h = H_{mac(k)}$ полученное из информации на втором этапе шифрования E''_k . Размер хеша определяется криптографической хеш-функцией SHA-256, т.е. 32 байтами. Таким образом, одно принятое, либо отправленное уникальное сообщение будет увеличивать БД на 32 байта. Период генерации одного сообщения одним узлом равен 5 секундам. Если предположить, что в сети существует 10 узлов, каждый из которых в период равный 5 секундам генерирует одно шифрованное сообщение, тогда за 5 секунд будет сгенерировано и сохранено 320 байт, либо 64 байт в секунду. Далее, если умножить полученный результат на количество секунд в минуте, минут в часе, часов в дне, дней в неделе, недель в месяце + 2 (30 дней), месяцев в году, то получим, что за год 10-ью узлами БД будет увеличена на $\sim 1.73\text{GiB}^4$ информации, что по современным меркам является вполне допустимым значением,

10. База данных хешей не защищает от воспроизведения злоумышленником старых сообщений, которые не были приняты узлом непосредственно владеющим БД. Такое событие становится возможным, если узел отсутствовал в периоды генерации шифртекстов оставшимися участниками системы. Вследствие этого, злоумышленник может нагрузить новоприбывший узел старыми и давно забытыми сообщениями сети. Защититься от этой атаки возможно, если подключиться к давно работающему анонимизирующему узлу или ретранслятору, сохранившему историю принятых сообщений. В таком случае он начнёт выступать в роли межсетевого экрана, разграничивая старые и новые шифртексты. Ещё одним возможным решением может стать обновление ключа сети k , если в системе было сгенерировано много трафика, а давно функционирующие узлы неизвестны или недоступны.

Добавление же метки времени t на каждое сообщение второго этапа шифрования напротив, способно нарушить анонимность QV-задачи, т.к. может создать дополнительную связь в генерации шифртекстов и сделать задачу систематичности $P_{(s)}$ легко решаемой. Так например, при создании истинного сообщения метка времени будет фиксироваться не на моменте отправления, а на моменте генерации. По причине того, что сообщения из очереди отправляются последовательно в сеть, то сгенерированный заранее ложный шифртекст

⁴ $(64 [\text{байт}] * 60 [\text{секунд}] * 60 [\text{минут}] * 24 [\text{часов}] * 7 [\text{дней}] * 4 [\text{недель}] + 2 [\text{дня}]) * 12 [\text{месяцев}] / 1024^3 [\text{GiB}] = 1.7303467020392418 \sim 1.73\text{GiB}$

будет обладать меткой времени приближенной к метке открытого сообщения. Иными словами, будет прослеживаться связь генерации истинных сообщений посредством существования двух меток времени, разница между которыми будет меньше одного периода. Ситуацию можно изменить, заменив ложный шифртекст истинным, а не просто отодвинув ложный шифртекст в очереди. Но, помимо данной проблемы, в механизме генерации шифртекстов должен быть также учтён момент указания метки времени t в зависимости от периода генерации T . Иными словами, для n -ого сообщения в очереди метка времени должна иметь следующий вид: $t = now + nT$, где now – текущее время. Данное решение может обладать также и негативным эффектом, если скорость генерации шифртекстов, например при большой загрузке узла, не будет поспевать за меткой времени. В таком случае может возникнуть рассинхронизация очередей по метке времени между ложными и истинными шифртекстами. Ситуацию можно изменить, постоянно удаляя старые сгенерированные шифртексты из очереди, но в таком случае могут теряться истинные сообщения, ещё даже не отправляясь в сеть, а также усложнится сама работа очереди, из-за чего могут возникать программные ошибки и уязвимости,

11. Сообщения в первом этапе шифрования имеют статичный размер равный 8192 байт, из которых 1658 байт уделяется заголовочным данным: вектор инициализации (16В), хеш (32В), подпись (512В), публичный ключ (526В), соль (32В), ключ шифрования (512В), размеры в байтах: хеша, подписи, публичного ключа, соли, ключа шифрования, шифрованного блока, данных, полезной нагрузки (28В). Второй этап шифрования добавляет минимум 68 байт (при отсутствии пустых байт), из которых: вектор инициализации (16В), доказательство работы (8В), хеш (32В), маска (4В), количество пустых байт (4В), размер сообщения после первого этапа шифрования (4В). Размер сообщения после полного шифрования становится равен 8260 байт из которых 6534 байт являются полезной нагрузкой. Таким образом, если период генерации сообщений равен 5 секундам, тогда за одну секунду сеть может передать ~1306 значимых байт. Если прикладное приложение предполагает коммуникацию типа «запрос-ответ», тогда вследствие произведённого запроса неминуемо начнётся этап ожидания ответа, что приведёт к двойному уменьшению пропускной способности до ~653 значимых байт в секунду из-за увеличенного интервала ожидания равного ~10 секундам. Итого, если будет присутствовать задача передачи файла размером 1MiB по сети, то передача файла будет занимать приблизительно от ~13.4 до ~26.8 минут (~803 и ~1606 секунд соответственно),

12. В отличие от классического описания QV-сетей, в реализации проекта go-реер структура очередей Q представлена двумя очередями: очередью истинных Q_k и очередью ложных Q_r сообщений далее сливающихся в одну. Необходимость в двух очередях обуславливается нуждой в фоновой генерации ложных сообщений, чтобы при достижении периода t очередь Q_r преимущественно была непустой. В свою очередь такая нужда связана непосредственно с алгоритмом доказательства работы, который значительно замедляет шифрование сообщений, вследствие чего выставленный период генерации t может быть расширен. Хоть ситуация в таком случае и схожа с *Примером 3.3*, тем не менее, систематичность алгоритма здесь не нарушается, т.к. генерация истинных и ложных сообщений происходит не на моменте их отправления, а на моменте помещения их в очередь,

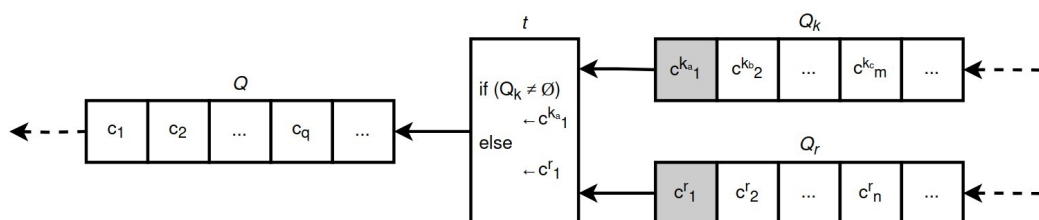


Рисунок 8. Схема двойной очереди сообщений в проекте go-peer

13. Анонимная сеть Hidden Lake может быть уязвима к активным кооперирующим наблюдателям, если таковые будут присутствовать в списке друзей у нескольких абонентов системы. В таком случае, помимо ухудшения качества анонимности – с задачи скрытия факта коммуникации до сокрытия связи между абонентами коммуникации, также появится и риск решения / деанонимизации самой задачи. Проблему можно было бы решить созданием нескольких очередей $Q_{k1}, Q_{k2}, \dots, Q_{kn}$ для каждого отдельного абонента, но в таком случае увеличится общее время ответа в n раз. Другим способом решения данной проблемы может стать изменение модели угроз, в которой друг априори будет равен доверенному узлу, а потому не будет заинтересован и задействован в процедурах деанонимизации. На таком конструкте F2F-систем базируется сеть Hidden Lake,

14. Модель угроз анонимной сети Hidden Lake ограничивается исключительно защитой сетевых коммуникаций между её участниками, что, в свою очередь, предполагает отсутствие каких-либо дополнительных мер, действий и условий направленных на защиту приватных ключей, баз данных, конфигурационных файлов или взаимодействий сервисов между собой в условиях локальной среды исполнения. Вследствие этого, предпринимаемые меры для обеспечения безопасности локального окружения, в условиях существования чувствительной информации, должны быть возложены на более низкие уровни взаимодействия, как например: выставление прав доступа к файлам и процессам, изолирование программ посредством использования виртуального окружения, программное и аппаратное полнодисковое шифрование, ограничение физического доступа к аппаратным средствам для третьих лиц и т.д.,

15. Сеть Hidden Lake имеет несколько дополнительных настроек, в первую очередь направленных на её адаптацию в других системах. Так например, настройки `rand_queue_period_ms>0` и `rand_message_size_bytes>0` делают период генерации и размер передаваемых сообщений динамичным параметром, что позволяет скрывать закономерные особенности генерируемого трафика от блокировочных систем. Параметр `work_size_bits>0` позволяет задавать сложность генерации сообщений в сети, благодаря чему становится возможным адаптация либо в сторону уменьшения значения – для слабых систем, либо в сторону увеличения значения – в условиях существования нескольких систем.

6. Заключение

В ходе работы было разобрано функционирование анонимной сети Hidden Lake на основе её математических моделей, с дальнейшим указанием преимуществ и ограничений, следующих из QB-задачи. Были приведены основные сервисы сети, а также их сетевые взаимодействия друг с другом в парадигме микросервисной архитектуры. Был описан стек

протоколов GP/12, позволяющий сетям с теоретически доказуемой анонимностью становиться абстрактными. Были представлены критерии выбора алгоритмов шифрования, подписания, хеширования, их параметров и настроек в моменты проектирования и реализации сети.

Список литературы

1. Popescu, B., Crispo, B., Tanenbaum, A. Safe and Private Data Sharing with Turtle: Friends Team-Up and Beat the System [Электронный ресурс]. — Режим доступа: https://web.archive.org/web/20070316085325/http://www.turtle4privacy.org/documents/sec_prot04.pdf (дата обращения: 04.07.2024).
2. Коваленко, Г. Общая теория анонимных коммуникаций. Второе издание / Г. Коваленко. — [б. м.]: Издательские решения, 2023. - 208 с.
3. Шнайер, Б. Прикладная криптография. Протоколы, алгоритмы и исходные коды на языке C / Б. Шнайер. — СПб.: ООО «Альфа-книга», 2018. - 1040 с.
4. Алферов, А., Зубов, А., Кузьмин, А., Черемушкин, А. Основы криптографии: Учебное пособие / А. Алферов, А. Зубов, А. Кузьмин, А. Черемушкин. — М.: Гелиос АРВ, 2002. - 480 с.
5. Граймс, Р. Апокалипсис криптографии / Р. Граймс. — М.: ДМК Пресс, 2020. - 290 с
6. Шнайер, Б., Фергюсон, Н. Практическая криптография / Б. Шнайер, Н. Фергюсон. - М.: Издательский дом «Вильямс, 2005. - 420 с.
7. Таненбаум, Э., Уэзеролл, Д. Компьютерные сети / Э. Таненбаум, Д. Уэзеролл. — СПб.: Питер, 2017. - 960 с.
8. Perry, M. Securing the Tor Network [Электронный ресурс]. — Режим доступа: <https://www.blackhat.com/presentations/bh-usa-07/Perry/Whitepaper/bh-usa-07-perry-WP.pdf> (дата обращения: 04.07.2024).
9. Astolfi, F., Kroese, J., Oorschot, J. I2P - Invisible Internet Project [Электронный ресурс]. — Режим доступа: https://staas.home.xs4all.nl/t/swtr/documents/wt2015_i2p.pdf (дата обращения: 04.07.2024).
10. Danezis, G., Dingledine, R., Mathewson, N. Mixminion: Design of a Type III Anonymous Remailer Protocol [Электронный ресурс]. — Режим доступа: <https://web.archive.org/web/20170312061708/https://gnunet.org/sites/default/files/minion-design.pdf> (дата обращения: 04.07.2024).
11. Reiter, M., Rubin, A. Crowds: Anonymity for Web Transactions [Электронный ресурс]. — Режим доступа: https://www.cs.utexas.edu/~shmat/courses/cs395t_fall04/crowds.pdf (дата обращения: 04.07.2024).
12. Ишкуватов, С. Способ и алгоритм определения типа трафика в зашифрованном канале связи [Электронный ресурс]. — Режим доступа: <https://cyberleninka.ru/article/n/sposob-i-algoritm-opredeleniya-tipa-trafika-v-shifrovannom-kanale-svyazi> (дата обращения: 04.07.2024).
13. Накамото, С. Биткойн: система цифровой пиринговой наличности [Электронный ресурс]. — Режим доступа: https://bitcoin.org/files/bitcoin-paper/bitcoin_ru.pdf (дата обращения: 04.07.2024).
14. Хлебников, А. OpenSSL 3: ключ к тайнам криптографии / А. Хлебников. — М.: ДМК Пресс, 2023. - 300 с.
15. Lewis, J., Fowler, M. Microservices [Электронный ресурс]. — Режим доступа: <https://martinfowler.com/articles/microservices.html> (дата обращения: 07.09.2024).

16. Коваленко, Г. Анонимная P2P-сеть внутри централизованного HTTPS-сервера: вшиваем паразитный трафик всеми правдами и неправдами [Электронный ресурс]. — Режим доступа: <https://habr.com/ru/articles/800965/> (дата обращения: 07.09.2024).
17. Heaton, R. The Padding Oracle Attack [Электронный ресурс]. — Режим доступа: <https://robertheaton.com/2013/07/29/padding-oracle-attack/> (дата обращения: 04.07.2024).
18. Омассон, Ж. О криптографии всерьез. Практическое введение в современное шифрование / Ж. Омассон. — М.: ДМК Пресс, 2021. - 328 с.
19. Мао, В. Современная криптография: теория и практика / В. Мао. - М.: Издательский дом «Вильямс, 2005. - 768 с.
20. Донован, А., Керниган, Б. Язык программирования Go / А.А. Донован, Б.У. Керниган. — М.: ООО «И.Д. Вильямс», 2018. - 432 с.
21. Barker, E. Recommendation for Key Management [Электронный ресурс]. — Режим доступа: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-57pt1r5.pdf> (дата обращения: 23.07.2024).