

Комплексная аналитическая оценка надежности анонимной сети

06, июнь 2017

Ершов Н. Г., Рязанова Н. Ю.

УДК: 004.738.2

Россия, МГТУ им. Н.Э. Баумана

yershov.n@mail.ru

ryaz_nu@mail.ru

Введение

Не ослабевающая заинтересованность в конфиденциальном общении в глобальной сети является стимулом для развития и совершенствования анонимных сетей. Конфиденциальность в глобальной сети является частной задачей обеспечения конфиденциальности данных и информационной безопасности. Анонимные сети предназначены для сокрытия факта контактов в глобальной сети, которые могут быть, как длительными, так и разовыми пересылками сообщений. Чем сильнее заинтересованность пользователей в анонимности (конфиденциальности), тем более надежной должна быть анонимная сеть. Процесс выбора оптимальных параметров анонимной сети связан с возможностью оценить эти параметры с точки зрения обеспечения надежного сокрытия устанавливаемой связи. Для оценки используются метрики анонимности. Чем шире набор параметров, на которые влияют выбранные метрики анонимности, тем более надежной будет построенная анонимная сеть.

1. Факторы обеспечения анонимности

Анализ трафика позволяет установить факт контактов в сети, чем и пользуются злоумышленники. Для устранения возможности обнаружения конфиденциальных контактов необходимо, чтобы узлы сети, через которые проходит взаимодействие, во-первых, не имели доступ к информации об отправителе и получателе сообщения, во-вторых, должна быть исключена возможность подмены сообщения и обеспечена сохранность закрытого ключа, и, в-третьих, каждый узел должен иметь возможность создавать, передавать и буферизовать сообщения в сети [1]. Для реализации этих требований необходимо, чтобы в виртуальной сети присутствовали три базовых компонента: анонимные прокси-сервера, локальный прокси-сервер и публичные сервера [2].

Основу анонимной сети составляют анонимные прокси-сервера, которые являются распределенными по сети точками доступа, обеспечивающими функциональные возможности перемешивающего узла. Такие точки доступа должны обладать доверием, регуляр-

но поддерживаться и обслуживаться. Пользователь обращается к анонимной сети через локальный прокси – сервер, который представляет собой запущенную на компьютере пользователя программу (приложение или службу). Локальный прокси-сервер обеспечивает помимо фильтрации трафика всю работу анонимной сети: создание и поддержка тоннелей, перемешивание узлов, послойное шифрование, буферизацию. В связи с этим возникает ряд факторов риска [3].

При отправке сообщения автоматически определяется множество доверенных анонимных прокси-серверов, составляющих анонимный тоннель, и устанавливаются параметры его параметры. Все соединения с компьютерами в сети осуществляются через этот тоннель. Аналогичные действия по построению тоннеля выполняются также принимающей стороной для отправки ответа получателю. Узел отправителя может входить в состав анонимной сети и участвовать в перемешивании узлов, или находиться вне анонимной сети и работать с ней через шлюз. В обоих случаях задача построения тоннеля должна лежать на узле отправителя сообщения. Кроме того, на узле отправителя сообщения шифруются.

У каждого узла сети имеется набор сервисов. Сервисы имеют разные требования к пропускной способности, надежности передачи, безопасности и стабильности. Тем не менее они должны работать в общей инфраструктуре, формируя ее параметры. Инфраструктура во многом определяет топологию сети, а топология напрямую влияет на степень анонимности. К инфраструктуре предъявляются следующие требования:

1. Высокая скорость работы. Сеть должна быть в состоянии обработать пакеты с достаточной скоростью, учитывая, что обеспечение анонимности требует затратных по времени манипуляций с пакетами.

2. Изолированность. Любой узел в рамках общей инфраструктуры должен иметь уникальный псевдоним, обеспечивающий надежную адресацию.

3. Гибкость. Сеть должна обеспечивать работу с разными типами пакетов и допускать появление новых типов пакетов.

4. Масштабируемость. Необходимо, чтобы сеть можно было расширять, осуществляя минимум технологических и структурных изменений в ней.

Комплексное обеспечение анонимности, т.е. обеспечение конфиденциального общения в сети, состоит в сокрытии как топологии, так и содержания сообщений. Другими словами, анонимность должна обеспечиваться как на уровне узлов сети, так и на уровне сообщений [3].

Для анонимных сетей помимо обеспечения работы тоннеля для передачи сообщений, важно защитить сообщения от возможных действий злоумышленников при попытке определить отправителя, получателя или промежуточные узлы тоннеля. Шифрование данных напрямую влияет на такие метрики анонимности как идентифицируемость и несвязность. Энтропия в сети тем выше, чем больше сообщений, для которых трудно определить их роль: запрос это или ответ.

Анонимность на уровне узлов достигается путем создания туннеля, который генерируется случайным образом из доверенных анонимных прокси-серверов и остается неизменным некоторое время [3]. Сохранение туннеля на это время необходимо для сокращения накладных расходов. Его перестроение является длительной операцией, поскольку построение туннеля требует проверки пула перемешиваемых узлов на работоспособность, генерации случайного пути, сбора ключей для шифрования. Именно длительное существование туннеля является одним из ключевых факторов риска. Длина туннеля и вероятность наличия в туннеле хотя бы одного доверенного узла определяет вероятность ненаблюдаемости в сети.

2. Метрики анонимности

Развитие способов обеспечения анонимности выполняется на основе результатов анализа проведенных атак с целью раскрытия анонимности. Каждая новая реализация анонимной сети должна учитывать вновь появившиеся факторы риска и обеспечить защиту от них. Необходимо иметь способ предиктивной оценки, позволяющий оценить вероятность того, что данные могут быть перехвачены, а анонимность нарушена. Поскольку анонимная сеть не стационарна и ее узлы могут подключаться и отключаться, такая оценка должна производиться динамически в процессе работы анонимной сети, что безусловно затрудняет оценку.

С повышением уровня анонимности и информационной безопасности неизбежно растут накладные расходы при работе сети. Требования к скорости работы сети и к типам передаваемых данных накладывают ограничения на реализацию способов обеспечения анонимности. При разработке анонимных сетей важно оценить не только вероятность нарушения анонимности, но и определить необходимый и достаточный размер сети, количество пользователей и максимально необходимое число сообщений.

Для сравнения реализаций анонимных сетей необходимо определить критерии оценки достигнутой степени анонимности. Эти критерии должны отражать изменения в реализации сети и определяться показателями выбранной метрики [4]. Предлагаются следующие критерии анонимности [5]:

- 1) Размер анонимного множества.
- 2) Вероятность несвязанности.
- 3) Вероятность необнаружимости.
- 4) Степень идентифицируемости.

3. Размер анонимного множества

Под анонимным множеством подразумевается такой набор узлов, участвующих в построении анонимной сети, который обеспечит невозможность идентификации каждого конкретного узла при их перемешивании. Важнейшим параметром анонимного множества является его размер. Согласно требованиям безопасности, вероятность того, что из n узлов можно определить достоверно отправителя и получателя равна $1/n$ [6]. Но в анонимной

сети узлы перемешиваются и набор узлов, непосредственно участвующих в перемешивании определяет пул перемешивания. Эффективный размер пула перемешивания зависит от интенсивности передачи сообщений. Работа сети детерминируется на небольшие временные интервалы – раунды. В каждом раунде происходит обмен информацией между всеми узлами сети. В случае если количество одновременно передаваемых сообщений в анонимной сети ограничено, не переданные сообщения хранятся в буфере и передаются в следующем раунде. Наличие буферизации влияет и на размер пула перемешивания, и на вероятность деанонимизации.

Рассмотрим сеть, состоящую из нескольких узлов. Сообщения буферизуются на узле и передаются в сеть раундами. Каждый раунд в сеть отправляется ровно N сообщений из буфера, причем не обязательно напрямую от отправителя к получателю, так как сообщение может быть несколько раз отправлено на разные узлы сети, прежде чем оно будет доставлено получателю. В буфере каждого узла остается еще n сообщений для отправки. Вероятность, что конкретное сообщение попадет в N отправленных сообщений равна:

$$p = \frac{N}{N+n}, \quad (1)$$

где N - количество сообщений, передающихся за один раунд, n - количество сообщений в буфере.

А вероятность того, что сообщение, доставленное на конечный узел за k раундов, было отправлено в раунде с номером χ равна:

$$P_{round\chi} = \frac{N}{N+n} \left(1 - \frac{N}{N+n}\right)^{k+\chi}, P_{round_0} = \left(\frac{n}{N+n}\right)^k. \quad (2)$$

Для расчёта размера сети используется определение информационной энтропии Шеннона [6]. Неопределенность того, что узел является отправителем конкретного сообщения равна сумме неопределенности на каждом раунде отправки:

$$S = - \sum_{\chi=1}^k p_{\chi} \times \log(p_{\chi}) - p_0 \times \log(p_0), \quad (3)$$

где S - информационная энтропия, k - число раундов, p_{χ} - вероятность отправки сообщения в раунде χ , p_0 - вероятность отправки сообщения в нулевом раунде при инициализации сети.

Если подставить выражение (2) в (3), то получим зависимость энтропии от количества сообщений:

$$S = - \sum_{\chi=1}^k \left(\frac{N}{N+n} \left(\frac{n}{N+n}\right)^{k-\chi} \times \log \left(\frac{N}{N+n} \left(\frac{n}{N+n}\right)^{k-\chi} \right) \right) - \left(\frac{n}{N+n}\right)^k \times \log \left(\frac{n}{N+n}\right)^k, \quad (4)$$

В случае, если число раундов стремится к бесконечности, энтропия может быть интерпретирована, как число вопросов, которые нужно задать, чтобы определить узел, с которого отправлено сообщение. Очевидным следствием этого является невозможность определения факта контакта, если количество вопросов, которые необходимо задать, пре-

вышает число узлов. Поэтому формула (4) может использоваться для вычисления размера пула перемешивания. Эффективный размер пула определяется по формуле:

$$\lim_{k \rightarrow \infty} S = \left(1 + \frac{n}{N+n}\right) \times \log(N+n) - \frac{n}{N} \times \log(n) - \log(N), \quad (5)$$

где S – эффективный размер пула перемешивания, n – число сообщений в буфере узла (в случае, если сообщения буферизуются на узле до отправки), N – число сообщений, пересылаемых в каждом раунде обмена.

Для анонимных сетей, предназначенных для обмена сообщениями важным аспектом является время отклика. Поэтому в сетях такого рода неприменима буферизация сообщений, неизбежно приводящая к задержкам. График зависимости размера пула перемешивания от числа заданного числа сообщений без буферизации показан на рисунке 1.

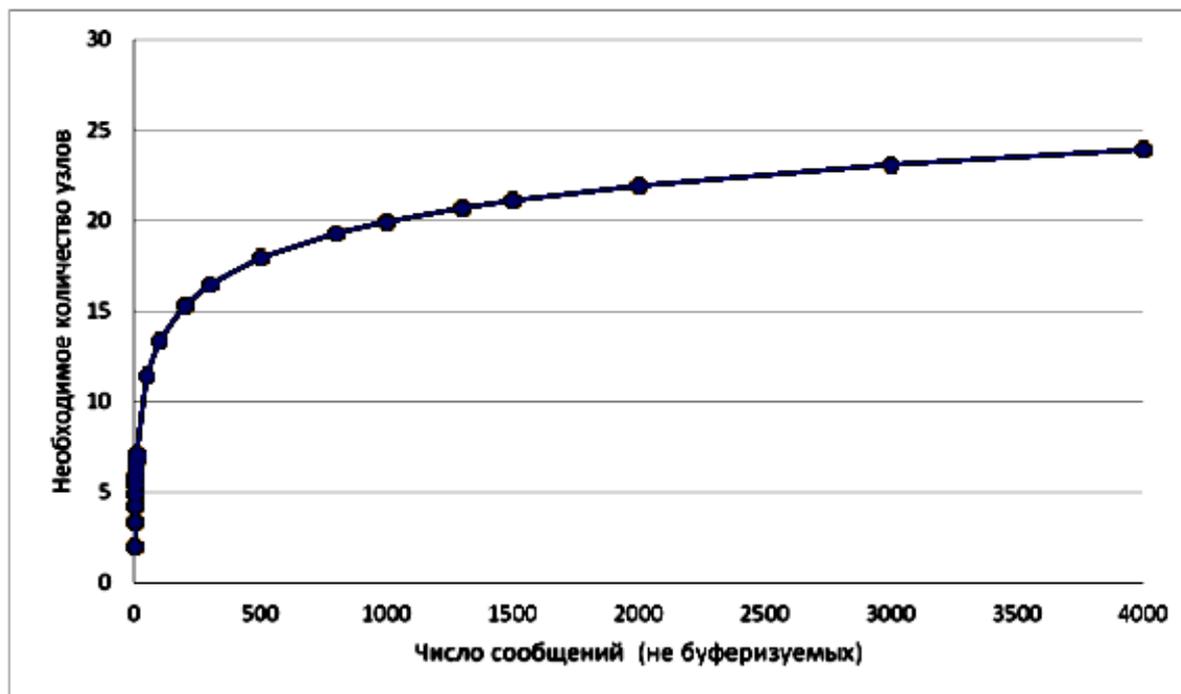


Рис. 1. График зависимости размера пула перемешивания от числа заданного числа сообщений без буферизации

Важнейшим фактором обеспечения анонимности, которой не уделено в литературе достаточного внимания, является буферизация. Буферизация является надежным средством ликвидации ряда факторов риска, но приводит к уменьшению числа сообщений в сети, а значит, к уменьшению энтропии.

Для того, чтобы энтропия оставалась на том же уровне необходимо искусственно увеличить число узлов в сети. График зависимости размера пула перемешивания от количества сообщений и размера буфера сообщений показан на рисунке 2.

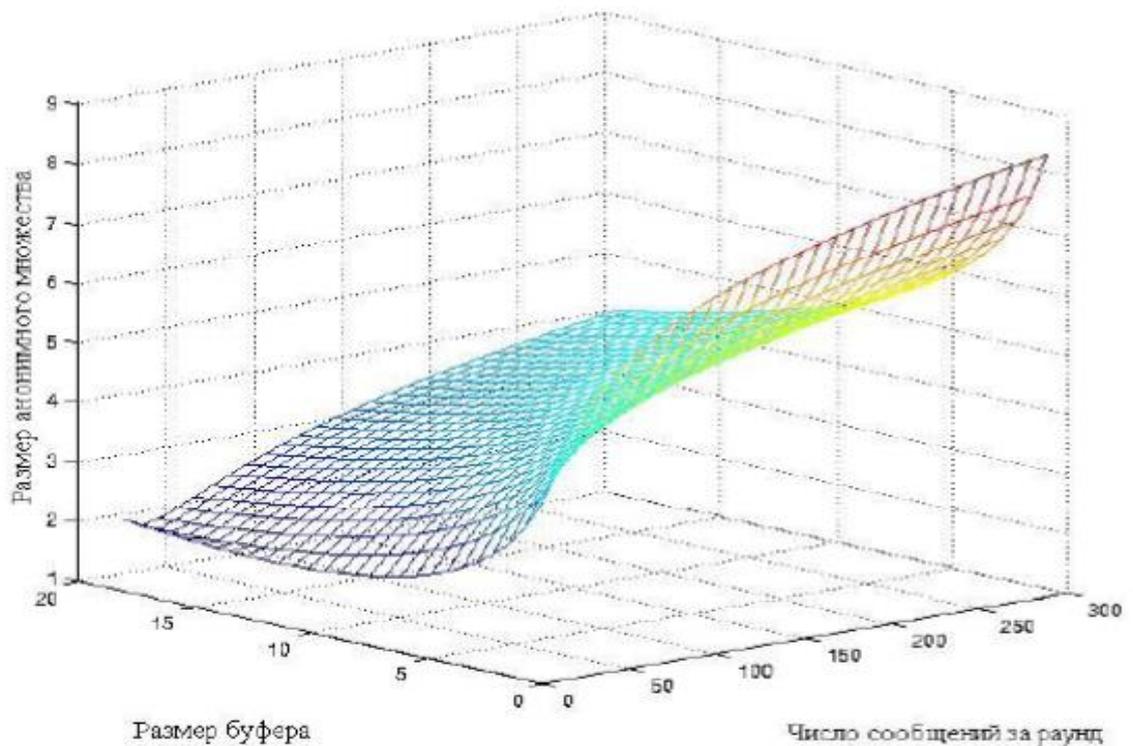


Рис. 2. График зависимости размера пула перемешивания от количества сообщений и размера буфера сообщений

4. Степень идентифицируемости

Под идентифицируемостью понимается возможность раскрытия факта контакта путем определения узла отправителя или получателя при передаче сообщения. Из всех узлов сети выбирается множество узлов, исполняющих роль отправителя или получателя. В этом множестве определяется энтропия – количество информации, необходимое для идентификации конкретного узла и его роли при текущей передаче сообщения:

$$I = -k \times \log_2(k) \quad (6)$$

где I - неопределенность факта контакта (энтропия);

k - вероятность того, что k -тый узел выполняет конкретную роль.

В передаче сообщений принимает участие пара узлов: отправитель и получатель, но таких пар в сети может быть множество. Отсюда количество информации, необходимое для их идентификации определяется, как увеличение энтропии [6]:

$$I(\psi) = -p_1 \times \log_2(p_1) - p_2 \times \log_2(p_2) - \dots - p_\psi \times \log_2(p_\psi) = -\sum_{i=1}^{\psi} p_i \times \log_2(p_i), \quad (7)$$

где ψ - мощность множества отправителей и получателей;

p_i - вероятность того, что i -тый узел выполняет конкретную роль.

Чем больше информации об отправителях и получателях имеется в распоряжении злоумышленника, тем меньше энтропия системы. Очевидно, если в пуле перемешивания

имеется только один узел, анонимность отсутствует. Показано, что максимально достижимая анонимность получается, когда $I = \log_2|\psi|$, [7]. Чем больше объем необходимой информации для раскрытия узла, тем надежнее анонимная сеть. График зависимости изменения показателя идентифицируемости от количества узлов, выполняющих роли отправителя и получателя представлен на рис. 3.

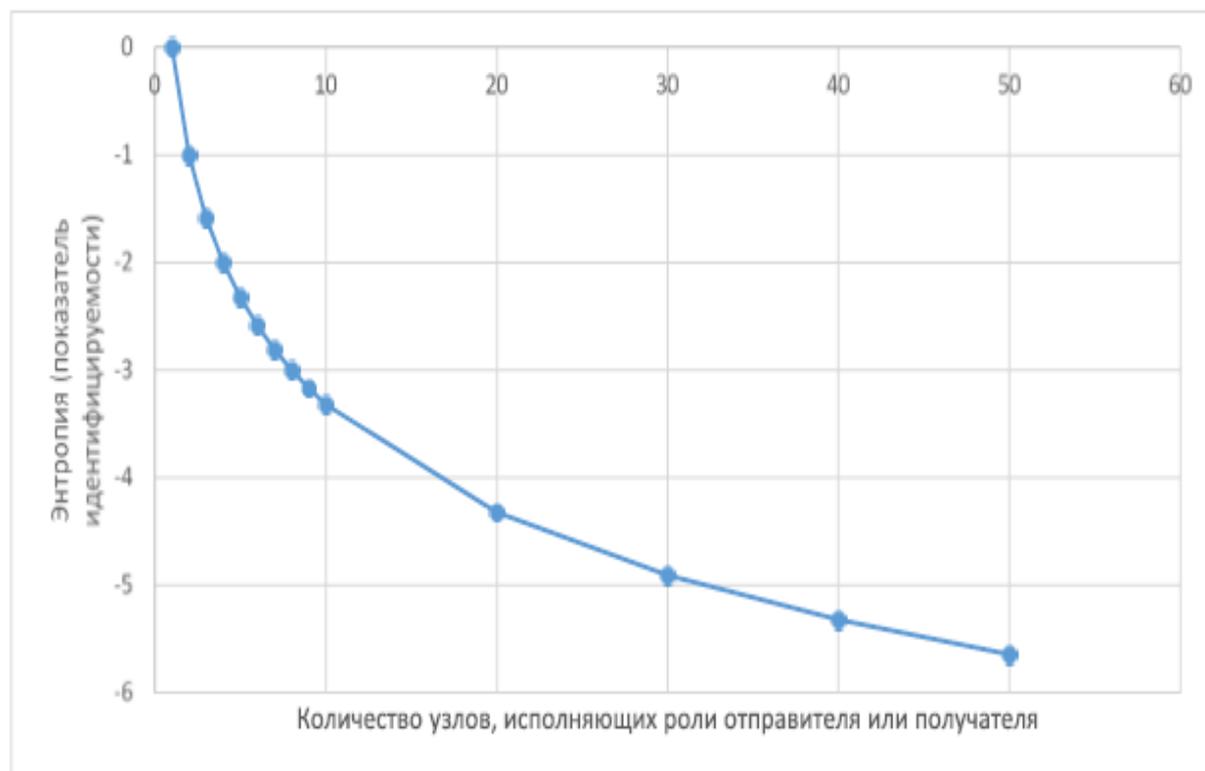


Рис. 3. График зависимости изменения показателя идентифицируемости от количества узлов, выполняющих роли отправителя и получателя

5. Вероятность ненаблюдаемости

Ненаблюдаемостью называется состояние сети, при котором сообщения в ней неотличимы друг от друга и невозможно определить истинность и ложность сообщения. Также при этом невозможно определить пару «отправитель – получатель» [5]. Делается пессимистичное предположение, что все транзитные узлы в анонимной сети являются потенциально атакующими. По условиям работы сети в ней должен находиться по крайней мере один доверительный узел (отправитель, входящий в состав пула перемешивания может являться доверительным узлом сам для себя). Схема обеспечения ненаблюдаемости после прохождения через доверительный узел представлена на рисунке 4.

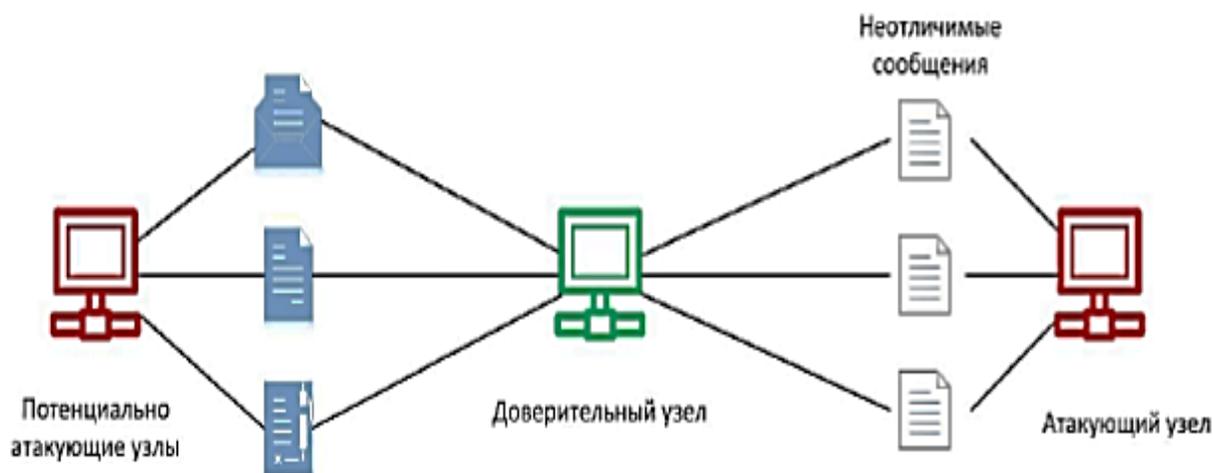


Рис. 4. Схема обеспечения ненаблюдаемости после прохождения через доверительный узел

Ненаблюдаемость оценивается вероятностью того, что по крайней мере на одном из узлов в пуле перемешивания, сообщения будут неотличимы друг от друга [3,8]. Поскольку неотличимость сообщений гарантируется только после прохождения через доверительный узел, определение ненаблюдаемости сводится к определению вероятности попадания доверительного узла в тоннель в качестве транзитного узла. Пусть необходимо построить тоннель некоторой длины из узлов, часть из которых является доверительными, а часть – потенциально атакующими. Вероятность попадания определяется из отношения количества тоннелей, проходящий через доверительный узел к количеству всевозможных тоннелей:

$$C_{M_A+M_G}^1 = \frac{(M_A+M_G)!}{(\ell!) \times (M_A+M_G-\ell)!}, \quad C_{M_A}^1 = \frac{M_A!}{\ell! \times (M_A-\ell)!}, \quad (8)$$

где $C_{M_A+M_G}^1$ - количество всевозможных тоннелей, M_G - количество доверительных узлов; M_A - количество потенциально атакующих узлов; ℓ - длина тоннеля, $C_{M_A}^1$ - количество тоннелей, в которых не входит ни один доверительный узел.

Вероятность того, что будет выбран тоннель, в который не входит ни один доверительный узел будет обратна вероятности ненаблюдаемости [9]:

$$p = 1 - \frac{C_{M_A}^1}{C_{M_A+M_G}^1} = 1 - \frac{M_A! \times (M_A+M_G-\ell)!}{(M_A-\ell)! \times (M_A+M_G)!}, \quad (9)$$

где M_A - количество атакующих узлов, M_G - количество доверительных узлов, ℓ - длина тоннеля или степень перемешивания.

Вероятность того, что среди перемешиваемых узлов будет хотя бы один доверительный узел, также зависит от интенсивности обмена сообщениями в сети и частотой перестроения туннеля перемешивания. Поскольку число узлов и длина тоннеля всегда положительны, то можно представить множители в формуле (9) используя первый член фор-

мулы Стирлинга и получить формулу для приближенного вычисления вероятности ненаблюдаемости:

$$p = 1 - \frac{M_A! \times (M_A + M_G - \ell)!}{(M_A - \ell)! \times (M_A + M_G)!} \approx 1 - \sqrt{\frac{M_A \times (M_A + M_G - \ell)}{(M_A - \ell) \times (M_A + M_G)}} \times \frac{M_A^{M_A} \times (M_A + M_G + \ell)^{M_A + M_G - \ell}}{(M_A - \ell)^{M_A - \ell} \times (M_A + M_G)^{M_A + M_G}} \quad (10)$$

График зависимости вероятности ненаблюдаемости от количества атакующих узлов на один доверительный узел для длины тоннеля равной 3 представлен на рисунке 5. Очевидно, что с ростом длины тоннеля вероятность прохождения сообщения через доверительный узел возрастает.

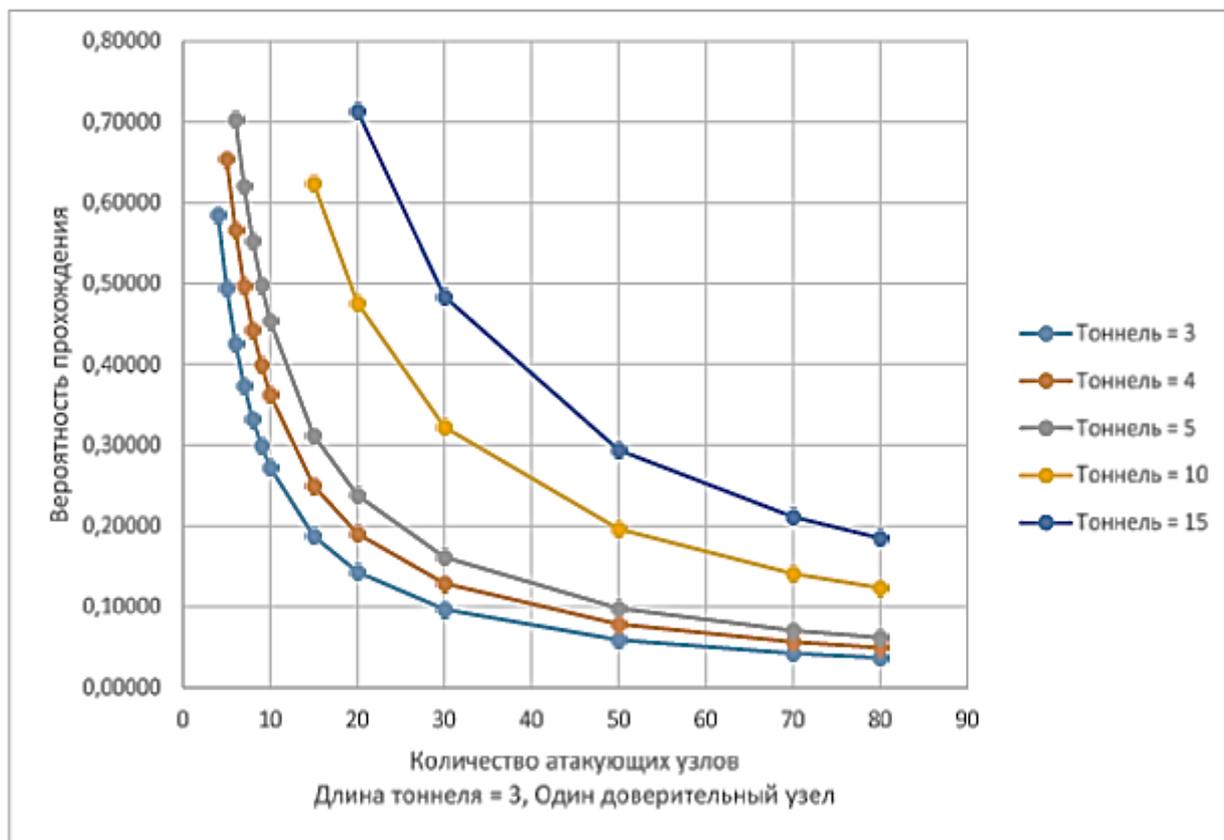


Рис. 5. График зависимости вероятности ненаблюдаемости от количества атакующих узлов на один доверительный узел для длины тоннеля равной 3

При достижении некоторой предельной величины рост количества атакующих узлов перестает существенно влиять и вероятность прохождения через доверительный узел нелинейно увеличивается.

Зависимость вероятности ненаблюдаемости от количества доверительных узлов и от количества атакующих узлов представлена на рисунке 6.

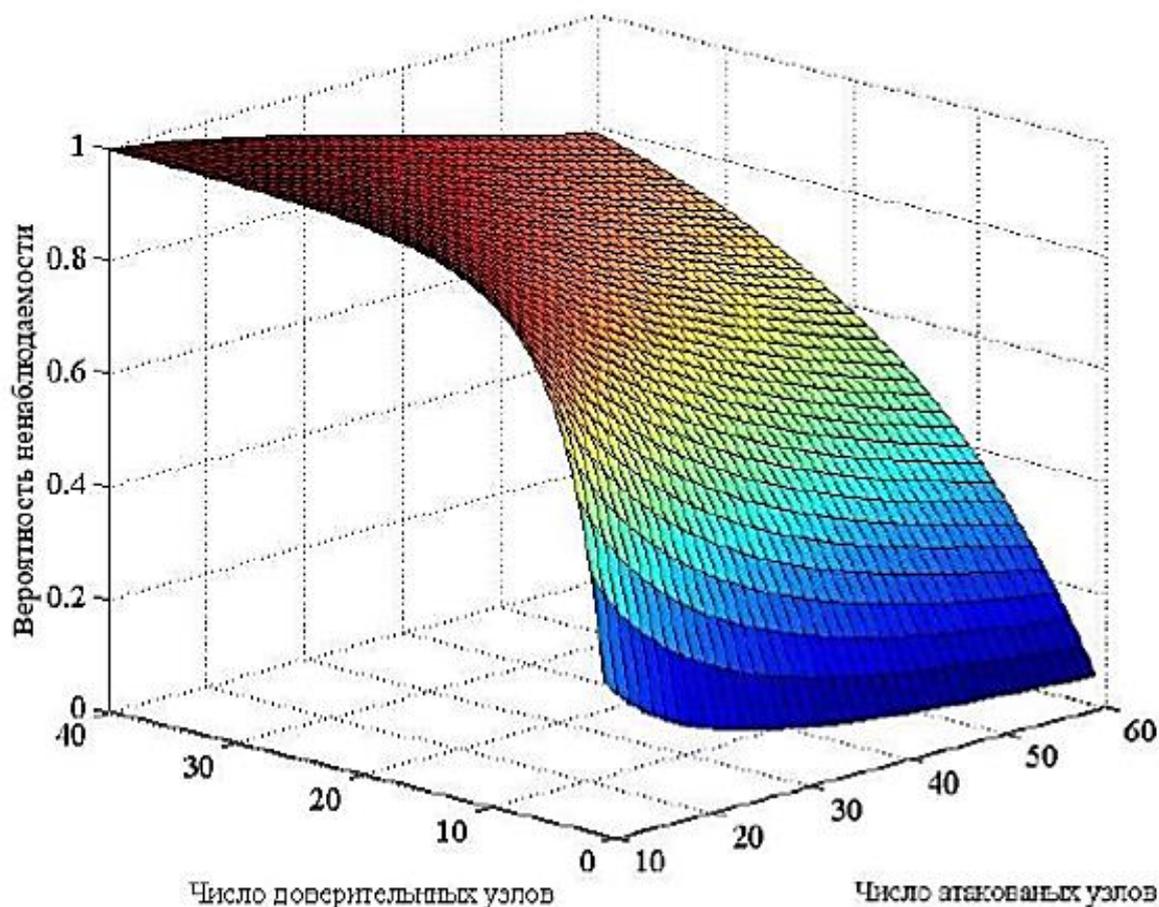


Рис. 6. Зависимость вероятности ненаблюдаемости от количества доверительных узлов и от количества атакующих узлов

Данная зависимость также нелинейна. Можно наблюдать, что при достаточно большом количестве доверительных узлов в сети даже значительное увеличение атакующих узлов не приводит к снижению вероятности ненаблюдаемости.

6. Вероятность несвязанности

Важной метрикой, позволяющей определить, трудоемкость установления связи аккаунта с конкретным узлом и сообщением является вероятность несвязанности. Под несвязанностью понимается невозможность определить соответствие между псевдонимами и их владельцами в процессе общения в анонимной сети.

Для определения несвязанности профилей, можно рассчитать вероятность того, что два перехваченных сообщения от разных псевдонимов относятся к одному реальному пользователю. Пусть для анонимной сети дана функция распределения количества сообщений в зависимости от аккаунта. Определим вероятность того, что два перехваченных сообщения принадлежат одному аккаунту [10]. Из определения вероятности следует:

$$P(d) = \frac{2F(d)}{\sum_{a \in D} F(a)}, \quad (11)$$

где $P(d)$ - вероятность принадлежности двух сообщений одному аккаунту;

d - аккаунт, которому принадлежит перехваченное сообщение;

$F(d)$ - количество сообщений, принадлежащих аккаунту d ;

D - множество зарегистрированных аккаунтов в сети.

Часто определить вероятность принадлежности двух сообщений одному аккаунту не представляется возможным, поскольку для этого требуются знания о текущем состоянии сети. В этой связи выдвигается предположение о такой вероятности и на основании этого вычисляется вероятность несвязанности. Исходя из формулы (11), вероятность того, что два перехваченных сообщения не принадлежат одному аккаунту равна: $P(d')=1-P(d)$. Если принять вероятность несвязанности μ , как вероятность верного определения узла отправителя, то ее можно будет вычислить как:

$$\mu \geq \frac{1}{r \times (1 - P(d))}, \quad (12)$$

где r - ожидаемое количество пользователей,

$P(d)$ - вероятность принадлежности сообщений одному аккаунту.

График зависимости вероятности несвязанности от количества аккаунтов и вероятности принадлежности сообщения к аккаунту представлена на рисунке 7.

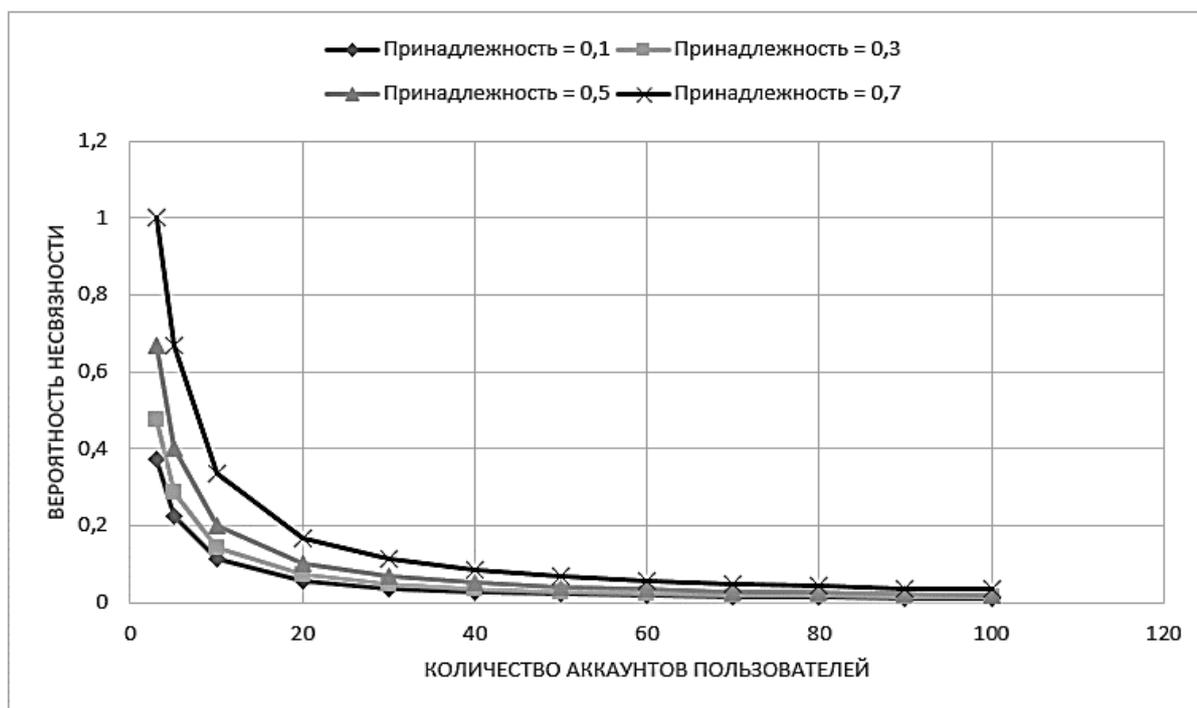


Рис. 7. График зависимости вероятности несвязанности от количества аккаунтов и вероятности принадлежности сообщения к аккаунту

Заключение

Для предиктивной оценки надежности работы анонимной сети и, следовательно, обеспечения конфиденциальности общения в виртуальной сети предлагается использовать

набор критериев анонимности: размер анонимного множества, вероятность несвязанности, вероятность необнаружимости, степень идентифицируемости.

Для каждой конкретной сети, на основании выбора факторов риска, от которых требуется реализовать защиту, можно определить степень ее надежности на основании предложенных аналитических зависимостей для выбранных метрик.

Список литературы

- [1]. Ершов Н.Г., Рязанова Н.Ю. Анализ типов атак на анонимную сеть и определение факторов риска // Сборник: Фундаментальные и прикладные исследования: проблемы и результаты труды II международной научно-практической конференции. 2016. С. 186-187.
- [2]. Marc Rennhard, Sandro Rafaeli, Laurent Mathy. Design, Implementation, and Analysis of an Anonymity Network for Web Browsing // Swiss Federal Institute of Technology, Computer Engineering and Networks Laboratory, Lancaster University, Faculty of Applied Sciences. Technical Report. 2002. №129
- [3]. Ершов Н.Г., Рязанова Н.Ю. Проблемы сокрытия трафика в анонимной сети и факторы, влияющие на анонимность // Инженерный журнал: наука и инновации. 2014. № 12 (36). С. 16.
- [4]. Andreas Pfitzmann, Marit Hansen. Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. A Consolidated Proposal for Terminology // Version v0.31, 2008
- [5]. Ершов Н.Г., Рязанова Н.Ю. Выбор метрик для оценки надежности сокрытия факта контакта в глобальной сети // Сборник: Теоретические и практические исследования XXI века труды II международной научно-практической конференции. 2016. С. 80-81.
- [6]. Добрушин Р.Л. Общая формулировка основной теоремы Шеннона в теории информации // УМН. 1959. № 14:6(90). С. 3–104
- [7]. Andrei Serjantov, George Danezis. Towards an Information Theoretic Metric for Anonymity // Proceedings of the 2Nd International Conference on Privacy Enhancing Technologies. 2003. Pp. 41-53
- [8]. Jian Ren, Jie Wub. Survey on anonymous communications in computer networks // Michigan State University, Temple University, USA, 2009
- [9]. Oliver Berthold, Andreas Pfitzmann, Ronny Standtke. The Disadvantages of Free MIX Routes and How to Overcome Them // Designing privacy enhancing technologies: proceedings / International Workshop on Design Issues in Anonymity and Unobservability, Berkeley, CA, USA. 2000. Pp. 30-45
- [10]. Peter Dornbach, Zoltan Nemeth. Privacy Enhancing Profile Disclosure // Privacy Enhancing Technologies, Second International Workshop, San Francisco, CA, USA, 2002. Pp. 85-98