UDC 004.056.53

Petriankin A.V., Grodzenskiy Ya.S. Identity provider as a single point of compromise in zero trust network access architectures

Провайдер идентификационных данных как единая точка компрометации в архитектурах доступа с нулевым доверием

Petriankin A.V.,
Master Student, Institute of Cyber Intelligence Systems,
National Research Nuclear University «MEPhI», Moscow, Russia
Grodzenskiy Ya.S.,
Candidate of technical sciences, Associate professor,
Institute of Cyber Intelligence Systems, National Research Nuclear University «MEPhI»,
Moscow, Russia
Петрянкин Алексей Владимирович,
магистрант, Институт интеллектуальных кибернетических систем, Национальный исследовательский ядерный университет «МИФИ» Россия, г. Москва
Гродзенский Яков Сергеевич,
кандидат технических наук,
Доцент кафедры «Криптология и кибербезопасность»
Национальный исследовательский ядерный университет «МИФИ»
Россия, г. Москва

Abstract. *Relevance. Zero Trust Network Access architectures reduce the network exposure of protected services, but move the center of trust to the identity infrastructure. An identity provider authenticates subjects, issues access tokens, transfers attribute assertions, and integrates with access policies. Its compromise, including administrative accounts, signing keys, or federated connections, can therefore affect many relying services. Objective. To develop an attack-surface model for an identity provider in ZTNA and classify the main compromise vectors of the identity plane. Methods. The study uses the NIST SP 800-207 architecture, set-theoretic attack-surface modeling, analysis of SAML 2.0, OpenID Connect and OAuth 2.0, and publicly described identity incidents and techniques. Results. The identity-plane attack surface is defined as a component of the ZTNA control-plane attack surface. The model distinguishes compromise of a user session, a signing key, administrative access, and federated trust. It shows that ZTNA reduces network exposure in normal operation, whereas IdP compromise produces a trust-concentration effect whose impact depends on the number and criticality of relying services. Scientific novelty. The novelty lies in modeling the IdP not merely as an authentication component but as an independent attack-surface object within the ZTNA control plane. Unlike approaches that focus primarily on network reachability in the data plane, the proposed model accounts for trust-violation scenarios: compromise of a user session, theft of signing keys, compromise of an IdP administrator, and breach of federated trust relationships. This makes it possible to quantify risk concentration in the identity infrastructure and compare architectures not only by the number of active network sessions but also by the amount of trust delegated to the identity provider.*

Keywords *and phrases: zero trust network access, ZTNA, identity provider, IdP, attack surface, identity plane, trust concentration, SAML, OpenID Connect, multi-factor authentication bypass.*

Аннотация. *Архитектуры доступа с нулевым доверием (Zero Trust Network Access, ZTNA) уменьшают сетевую экспонированность защищаемых сервисов: пользователь получает доступ не к внутреннему сегменту сети, а к конкретному приложению или ресурсу. Однако при этом центр доверия переносится с сетевого периметра на инфраструктуру идентификации. Провайдер идентификационных данных (Identity Provider, IdP) становится узлом, через который проходят аутентификация, выпуск маркеров доступа, передача атрибутных утверждений и интеграция с политиками доступа. Поэтому компрометация IdP, его административных учетных записей, ключей подписи или федеративных связей способна затронуть сразу множество сервисов, доверяющих его утверждениям. Цель: разработка модели оценки поверхности атаки провайдера идентификационных данных в*

ZTNA-архитектурах и классификация основных векторов компрометации плоскости идентификации. Методы. Использованы архитектурный анализ модели Zero Trust по NIST SP 800-207, теоретико-множественное моделирование поверхности атаки, анализ протоколов и стандартов SAML 2.0, OpenID Connect и OAuth 2.0, а также обобщение публично описанных инцидентов и техник атак на инфраструктуру идентификации. Результаты: введено понятие поверхности атаки плоскости идентификации как частной составляющей поверхности атаки плоскости управления ZTNA. Предложена сценарная модель экспонированности IdP, различающая компрометацию пользовательской сессии, ключа подписи, административного доступа и федеративного доверия. Показано, что в штатном режиме ZTNA может сокращать сетевую экспонированность сервисов, но при компрометации IdP возникает эффект концентрации доверия, при котором масштаб последствий определяется числом сервисов, доверяющих данному провайдеру идентификационных данных. Научная новизна и элемент авторского вклада состоят в том, что IdP рассматривается не только как компонент аутентификации, а как самостоятельный объект моделирования поверхности атаки плоскости управления ZTNA. В отличие от подходов, оценивающих преимущественно сетевую достижимость сервисов в плоскости передачи данных, предложенная модель учитывает сценарии нарушения доверия: захват пользовательской сессии, кражу ключей подписи, компрометацию администратора IdP и нарушение федеративных связей. Это позволяет количественно описывать концентрацию риска в инфраструктуре идентификации и сопоставлять архитектуры не только по числу активных сетевых сессий, но и по объему доверия, делегированного IdP.

***Ключевые слова**: доступ с нулевым доверием, ZTNA, провайдер идентификационных данных, IdP, поверхность атаки, плоскость идентификации, единая точка компрометации, SAML, OpenID Connect, обход многофакторной аутентификации.*

Рецензент: Сагитов Рамиль Фаргатович - кандидат технических наук, доцент. Заместитель директора, главный научный сотрудник. ООО «Научно-исследовательский проектный институт «Промышленное и гражданское строительство»

Развитие распределенных информационных систем привело к тому, что традиционная модель защищенного сетевого периметра перестала соответствовать практике эксплуатации корпоративных ресурсов. Пользователи, устройства и приложения находятся в разных сетях, облачных средах и организационных контурах. В этих условиях архитектуры доступа с нулевым доверием предлагают иной подход: доступ предоставляется не к сети в целом, а к конкретному ресурсу на основании идентичности субъекта, состояния устройства, контекста запроса и действующей политики безопасности. [1]

В стандарте NIST SP 800-207 выделяются логические компоненты архитектуры Zero Trust: модуль принятия решений о доступе, администратор политик и точка применения политик. На практике эти компоненты тесно связаны с системой управления идентификацией и доступом, центральным элементом которой становится провайдер идентификационных данных. Он подтверждает личность пользователя или сервиса, выдает маркеры доступа, передает атрибутные утверждения о группах и ролях и участвует в оценке риска сессии. [2]

Такое смещение повышает управляемость доступа, но формирует новое архитектурное противоречие. С одной стороны, ZTNA уменьшает сетевую поверхность атаки: сервисы становятся недоступны без авторизации, а пользователь не получает

широкую связность с внутренним сегментом. С другой стороны, множество доверяющих сервисов начинают принимать решения на основании утверждений одного IdP. Если нарушитель получает возможность выпускать действительные маркеры доступа, изменять политики, управлять многофакторной аутентификацией или подменять федеративные отношения, компрометация распространяется уже не через сеть, а через доверие.

Поэтому для оценки безопасности ZTNA недостаточно анализировать только плоскость передачи данных. Необходимо отдельно рассматривать плоскость управления и, внутри нее, плоскость идентификации. Настоящая статья направлена на формализацию этого аспекта: провайдер идентификационных данных рассматривается как потенциальная точка концентрации компрометационного риска, а не только как вспомогательный сервис аутентификации. В существующих работах по Zero Trust основное внимание уделяется принципам непрерывной проверки доступа, микросегментации и снижению сетевой достижимости сервисов. Отдельные исследования поверхности атаки описывают общие метрики экспонированности, а материалы по SAML, OpenID Connect и OAuth 2.0 фиксируют конкретные техники атак на инфраструктуру идентификации. Вместе с тем недостаточно явно рассматривается провайдер идентификационных данных как самостоятельный объект концентрации доверия в плоскости управления ZTNA. Именно этот аспект рассматривается в настоящей работе.

В типовой ZTNA-системе IdP участвует в следующих процессах: аутентификация субъекта: подтверждение личности пользователя, устройства, сервисной учетной записи или программного компонента; формирование атрибутных утверждений: передача сведений о ролях, группах, контексте, состоянии устройства, факте прохождения многофакторной аутентификации и уровне риска; выдача маркеров доступа: формирование SAML-утверждений, маркеров доступа, маркеров обновления и иных артефактов, используемых доверяющими сервисами; интеграция с политиками доступа: передача атрибутов в модуль принятия решений для разрешения или запрета конкретной сессии; федерация доверия: взаимодействие с внешними провайдерами идентификационных данных, партнерами, облачными платформами и прикладными сервисами.

Наиболее распространенными механизмами взаимодействия являются SAML 2.0, OpenID Connect и OAuth 2.0. При этом SAML 2.0 и OpenID Connect используются для передачи утверждений об аутентификации и идентичности, тогда как OAuth 2.0 является протоколом авторизации и применяется для выдачи и использования маркеров доступа. В SAML-модели IdP подписывает утверждение, которое принимает доверяющий сервис.

В модели OpenID Connect, построенной поверх OAuth 2.0, провайдер идентификационных данных или сервер авторизации выдает подписанные маркеры, содержащие набор атрибутов. В обоих случаях сервис принимает решение не на основании самостоятельной повторной проверки пользователя, а на основании доверия к IdP и его криптографическим ключам. [3, 4]

Критическим свойством такой модели является асимметрия доверия. Доверяющий сервис, как правило, не выполняет независимую повторную проверку субъекта, а принимает утверждение IdP. Следовательно, компрометация IdP, его ключей подписи или административной конфигурации может приводить к выпуску действительных для сервисов артефактов доступа.

Термин «единая точка отказа» описывает компонент, отказ которого приводит к нарушению работы всей системы. Для провайдера идентификационных данных в ZTNA более точным является термин «потенциальная точка концентрации компрометационного риска». Под ним в настоящей работе понимается компонент, нарушение доверия к которому при определенных условиях централизации создает возможность несанкционированного доступа к множеству зависимых сервисов. [5, 6]

IdP может становиться потенциальной точкой концентрации компрометационного риска по следующим причинам: множество доверяющих сервисов принимают утверждения одного источника идентификационных данных; административные действия в IdP могут менять доступ сразу к нескольким приложениям; компрометация ключей подписи позволяет формировать криптографически действительные маркеры или SAML-утверждения; отзыв доверия к IdP в рабочем режиме затруднен, поскольку приводит к нарушению доступа пользователей к критичным сервисам; федеративные связи расширяют границы доверия за пределы одной организации.

Важно подчеркнуть, что этот вывод не отрицает преимуществ ZTNA. Архитектура ZTNA действительно может уменьшать поверхность атаки в плоскости передачи данных. Однако выигрыш в плоскости передачи данных сопровождается ростом значимости IdP как компонента плоскости управления. Следовательно, корректная оценка безопасности должна учитывать обе составляющие.

Полную поверхность атаки ZTNA-системы целесообразно представить как сумму трех компонент:

$$AS_{total}(t) = AS_{data}(t) + AS_{control}(t) + AS_{endpoint}(t) \quad (1)$$

где $AS_{data}(t)$ — экспонированность плоскости передачи данных; $AS_{control}(t)$ — экспонированность плоскости управления, включая IdP, контроллер политик, каналы

доставки решений и механизмы отзыва; $AS_{endpoint}(t)$ — уязвимость конечных устройств, агентов и локальных хранилищ ключей.

Предметом настоящей статьи является частная компонента:

$$AS_{idp}(t) \subset AS_{control}(t)$$

то есть поверхность атаки плоскости идентификации.

Пусть U — множество пользователей и сервисных субъектов; S — множество сервисов, доверяющих IdP; A — множество административных учетных записей IdP; F — множество федеративных связей; K — множество ключей подписи и криптографических секретов IdP; w_s — вес сервиса s, отражающий его критичность и объем предоставляемого доступа. Через $K_t \subseteq U \times S$ обозначим множество активных направлений доступа в момент времени t, то есть множество пар «субъект — сервис», для которых существует хотя бы одна действующая сессия или действительный маркер доступа. Если между одним субъектом и одним сервисом одновременно существует несколько технических сессий, в базовой модели они агрегируются в одну пару (u, s), поскольку оценивается не число соединений, а охват доверия и достижимость сервиса.

Для сценария компрометации пользовательской учетной записи или сессии поверхность атаки можно определить следующим образом:

$$AS_{cred}^{idp}(t) = \sum_{(u,s)\in K_t} w_s \cdot m_u \cdot e_s \quad (2)$$

где m_u — коэффициент уязвимости аутентификации пользователя u, а e_s — коэффициент окна эксплуатации маркера доступа для сервиса s. Значение m_u снижается при использовании устойчивой к фишингу многофакторной аутентификации, привязки к устройству и проверки состояния конечной точки. Значение e_s снижается при коротком сроке действия маркеров, наличии механизма отзыва и непрерывной оценке доступа. [7, 8]

Компрометация IdP не всегда ограничивается активными сессиями. Для сценария кражи ключа подписи SAML или OpenID Connect поверхность атаки определяется уже не текущими сессиями, а всем множеством сервисов, доверяющих данному ключу: [9]

$$AS_{key}^{idp} = \chi_{key} \sum_{s\in S_{key}} w_s \quad (3)$$

где χ_{key} — индикатор компрометации ключа, а S_{key} — множество сервисов, принимающих маркеры или SAML-утверждения, подписанные этим ключом. При $\chi_{key} = 1$ нарушитель получает возможность формировать действительные артефакты доступа в обход обычного процесса аутентификации.

Для сценария компрометации административной учетной записи предлагается выражение:

$$AS_{admin}^{idp} = \sum_{a \in A_{comp}} \quad \sum_{s \in S(a)} \quad w_s \text{ (4)}$$

где A_{comp} — множество скомпрометированных администраторов, а $S(a)$ — множество сервисов, на которые администратор может влиять через изменение политик, групп, факторов аутентификации, приложений или доверенных сертификатов.

Для сценария компрометации федеративного партнера:

$$AS_{fed}^{idp} = \sum_{f \in F_{comp}} \quad \sum_{s \in S(f)} \quad w_s \text{ (5)}$$

где F_{comp} — множество нарушенных федеративных связей, а $S(f)$ — сервисы, принимающие утверждения через данную федерацию.

Обобщенный коэффициент концентрации доверия IdP может быть представлен как:

$$ICF = max\left(AS_{cred}^{idp}, AS_{key}^{idp}, AS_{admin}^{idp}, AS_{fed}^{idp}\right) \text{ (6)}$$

Использование максимума отражает сценарную, а не вероятностную природу показателя. В рамках одного расчетного сценария выбирается доминирующий способ первичной компрометации: пользовательская сессия, ключ подписи, административный доступ или федеративное доверие. Суммирование этих компонент в базовой модели не применяется, поскольку один и тот же сервис может попадать сразу в несколько множеств достижимости, что приведет к двойному счету. Для анализа сложного инцидента, в котором несколько сценариев реализуются последовательно или совместно, может быть введен отдельный комбинированный сценарий либо взвешенная сумма компонент; такая вероятностная оценка выходит за рамки настоящей статьи.

Предложенная модель не является вероятностной оценкой риска. Она не отвечает на вопрос, насколько вероятна компрометация. Ее назначение — определить потенциальный охват компрометации при реализации конкретного сценария и тем самым количественно описать концентрацию доверия в IdP.

Для демонстрации работоспособности модели рассмотрим условную организацию, в которой через один IdP доступны четыре сервиса: корпоративная почта s_1 с весом $w_{s_1} = 3$, репозиторий исходного кода s_2 с весом $w_{s_2} = 5$, система непрерывной интеграции s_3 с весом $w_{s_3} = 4$ и кадровая система s_4 с весом $w_{s_4} = 2$. Вес задается экспертно по шкале от 1 до 5, где 1 соответствует низкой критичности, а 5 — сервису, компрометация которого дает значимый ущерб или возможность дальнейшего развития атаки.

Пусть пользователь u_1 имеет активные направления доступа только к почте и репозиторию. Для него используется устойчивая к фишингу многофакторная аутентификация и проверка устройства, поэтому принимаем $m_{u_1} = 0{,}2$. Для почты срок

действия маркера и механизм отзыва дают $e_{s_1} = 0{,}2$, для репозитория — $e_{s_2} = 0{,}3$. Тогда пользовательский сценарий дает:

$$AS_{cred}^{idp}(t) = 3 \cdot 0{,}2 \cdot 0{,}2 + 5 \cdot 0{,}2 \cdot 0{,}3 = 0{,}42.$$

Если же скомпрометирован ключ подписи, которому доверяют все четыре сервиса, то охват определяется не активными сессиями, а всем множеством доверяющих сервисов:

$$AS_{key}^{idp} = 3 + 5 + 4 + 2 = 14.$$

Если скомпрометирован оператор службы поддержки, способный управлять доступом только к почте и кадровой системе, то:

$$AS_{admin}^{idp} = 3 + 2 = 5.$$

Таблица 1

Связь сценариев компрометации с мерами защиты

Сценарий	Что расширяет охват	Что снижает охват	Практический контроль
Компрометация сессии	Длительные маркеры, отсутствие отзыва, слабая MFA	Короткие токены, отзыв, проверка устройства	Phishing-resistant MFA, CAE/отзыв, device binding
Кража ключа подписи	Один ключ для множества сервисов, экспортируемые ключи	Разделение ключей, HSM/KMS, ротация	Невыгружаемые ключи, журналирование подписи, emergency rotation
Компрометация администратора	Широкие права IdP-админов, общая учётная запись	JIT/JEA-доступ, разделение контуров	PAM, отдельные admin accounts, approval workflow
Компрометация федерации	Неограниченные federation trust и OAuth consent	Срок действия, scoped claims, периодический пересмотр	Реестр federation trust, владельцы, review, журналирование consent

В рассматриваемом примере $ICF = max(0{,}42; 14; 5; 0) = 14$, то есть доминирующим сценарием является компрометация ключа подписи. Пример показывает, что усиление пользовательской аутентификации снижает компоненту $AS_{cred}^{idp}(t)$, но не устраняет необходимость защиты криптографических ключей и административной плоскости IdP.

Значение $AS_{fed}^{idp} = 0$, поскольку в рассматриваемом примере федеративные связи не заданы.

Научная новизна состоит в следующем.

Во-первых, поверхность атаки ZTNA-архитектуры разделена на плоскость передачи данных, плоскость управления и плоскость конечных устройств, при этом плоскость идентификации выделена как самостоятельная часть плоскости управления.

Во-вторых, для IdP предложена сценарная модель экспонированности, которая различает четыре типа компрометации: пользовательскую, криптографическую, административную и федеративную. Такое разделение позволяет показать, что разные атаки имеют разный радиус действия: от одной активной сессии до полного множества сервисов, доверяющих IdP.

В-третьих, введен коэффициент концентрации доверия, позволяющий количественно описывать зависимость масштаба последствий от числа доверяющих сервисов, объема административных полномочий и количества федеративных связей.

Векторы атак на IdP целесообразно разделить на пять классов.

Таблица 2

Классификация векторов компрометации IdP

Класс атак	Примеры	Механизм	Последствия
Компрометация пользователя	подбор паролей, использование утекших паролей, фишинг, утомление запросами МФА	получение пароля, сессии или подтверждения второго фактора	доступ к сервисам в рамках прав пользователя
Компрометация сессии	атака «злоумышленник посередине», кража файлов сеанса, кража маркера обновления	перехват уже аутентифицированной сессии	обход МФА после успешной аутентификации
Компрометация криптографического доверия	кража ключа подписи, подмена сертификата подписи, атака Golden SAML	выпуск действительных утверждений или маркеров	имитация пользователей для всех доверяющих сервисов
Компрометация администрирования	захват учетной записи администратора, злоупотребление службой поддержки, сброс МФА	управление политиками и учетными записями	повышение привилегий, отключение защитных механизмов
Компрометация федерации и интеграций	подмена внешнего IdP, злоупотребление согласием приложения OAuth, злоупотребление SCIM/API	нарушение доверенных связей между доменами	распространение компрометации между организациями и сервисами

К первому классу относятся атаки на первичную аутентификацию: подбор паролей, использование утекших учетных данных, фишинг и социальная инженерия. Такие атаки обычно ограничены правами конкретного пользователя, но при компрометации привилегированной учетной записи их эффект становится системным. [10]

Ко второму классу относятся атаки на сессионные артефакты. Особенность этих атак состоит в том, что многофакторная аутентификация может быть уже успешно пройдена пользователем, а нарушитель получает не пароль, а действительную сессию

или маркер доступа. Это снижает эффективность традиционных способов МФА, не защищенных от фишинга и перехвата сессии. [11]

Третий класс является наиболее опасным с архитектурной точки зрения. При компрометации ключей подписи нарушитель может формировать маркеры или SAML-утверждения, которые доверяющий сервис воспринимает как легитимные. В этом случае атака выходит за пределы одной пользовательской сессии. [12]

Четвертый класс связан с административной плоскостью IdP. Нарушитель, получивший права администратора, может добавлять приложения, изменять группы, сбрасывать факторы аутентификации, создавать резервные учетные записи, менять политики доступа и регистрировать новые ключи. Такие действия часто выглядят как штатная административная активность и требуют отдельного контроля. [13]

Пятый класс относится к федеративным связям и интеграциям. Чем больше внешних IdP, прикладных сервисов, коннекторов синхронизации, OAuth-приложений и сервисных учетных записей подключено к IdP, тем шире область возможного распространения компрометации.

В классической VPN-архитектуре точкой концентрации риска является VPN-шлюз. Его компрометация открывает нарушителю сетевую достижимость внутренних ресурсов. В ZTNA сетевой доступ ограничивается, но роль точки концентрации риска переносится в плоскость идентификации.

Таблица 3

Сравнение VPN-шлюза и IdP как точек концентрации риска

Характеристика	VPN-шлюз	IdP в ZTNA
Плоскость архитектуры	плоскость передачи данных	плоскость управления и идентификации
Основной объект доверия	сетевой туннель	маркер доступа, утверждение, атрибут
Последствие компрометации	доступ к сетевому сегменту	доступ к сервисам, доверяющим IdP
Типовое развитие атаки	горизонтальное перемещение по сети	изменение ролей, групп, приложений и маркеров
Ключевой риск	широкая сетевая достижимость	концентрация доверия и ключей
Основные меры снижения	микросегментация, списки доступа, переход к ZTNA	сегментация IdP, защита ключей, МФА, управление привилегиями, мониторинг маркеров

Это сравнение показывает, что переход от VPN к ZTNA не устраняет проблему концентрации компрометационного риска автоматически. Он изменяет ее природу: вместо сетевого шлюза критическим объектом становится провайдер идентификационных данных и связанная с ним инфраструктура доверия.

Снижение риска компрометации IdP должно строиться не только на усилении аутентификации пользователей, но и на уменьшении концентрации доверия.

Сегментация доверия IdP. Критичные сервисы не должны безусловно доверять тому же IdP и тем же политикам, что и массовые корпоративные приложения. Для административных, финансовых и технологических систем целесообразно использовать отдельные зоны доверия, отдельные политики или дополнительный независимый фактор подтверждения.

Разделение пользовательского и административного контуров. Административный доступ к IdP должен быть отделен от обычного доступа единого входа. Администратор IdP не должен использовать одну и ту же учетную запись для повседневной работы и для управления политиками.

Минимизация федеративных связей. Каждая федеративная связь должна иметь владельца, срок действия, документированную цель, ограниченную область действия и процедуру регулярного пересмотра.

Ограничение согласий приложений и интеграций. Регистрация приложений, выдача согласия OAuth и создание сервисных учетных записей должны проходить через утверждение и журналироваться как привилегированные операции.

Хранение ключей подписи в аппаратном модуле или сервисе управления ключами. Ключи SAML и OpenID Connect должны быть неэкспортируемыми, а операции подписи — журналируемыми.

Регулярная ротация ключей и сертификатов. Ротация должна выполняться планово и внепланово после инцидентов, связанных с подозрением на компрометацию.

Короткий срок действия маркеров доступа. Маркер доступа должен иметь минимально допустимый срок действия, а маркер обновления — привязку к устройству, контексту и политике риска.

Поддержка механизма отзыва. Доверяющие сервисы должны учитывать отзыв маркеров, изменение риска и отключение пользователя без ожидания полного истечения срока действия маркера.

Устойчивая к фишингу многофакторная аутентификация. Для администраторов и пользователей критичных систем следует применять аппаратные ключи, WebAuthn или сертификатную аутентификацию. Подтверждения через уведомления и одноразовые коды менее устойчивы к социальной инженерии и перехвату сессии.

Управление привилегированным доступом. Административные сессии должны быть временными, подтверждаемыми, журналируемыми и ограниченными по принципу доступа «точно в срок».

Аварийные учетные записи. Резервные учетные записи должны существовать, но храниться отдельно, использовать аппаратные факторы и регулярно проверяться. Их применение должно сопровождаться немедленным уведомлением службы безопасности.

Ограничение операций службы поддержки. Сброс МФА, смена номера телефона, восстановление доступа и изменение факторов аутентификации должны считаться высокорисковыми операциями.

Для IdP необходимо выделить отдельный набор признаков возможной компрометации: массовая выдача маркеров без соответствующих событий интерактивного входа; входы администраторов с нетипичных устройств, сетей и географий; изменение SAML- или OpenID Connect-конфигурации, сертификатов и адресов перенаправления; регистрация новых OAuth-приложений с широкими правами; рост числа запросов МФА к одному пользователю; успешная аутентификация после серии отказов МФА; массовый экспорт пользователей, групп или конфигураций через API; изменение коннекторов синхронизации и сервисных учетных записей; появление маркеров с нетипичным сроком действия или набором атрибутов.

Реагирование на компрометацию IdP должно включать не только блокировку пользователя, но и ротацию ключей, отзыв маркеров, проверку приложений, анализ федеративных связей и пересмотр административных действий за период возможного присутствия нарушителя.

Предложенная модель оценивает потенциальный охват компрометации, но не вычисляет вероятность реализации атаки. Для полноценной оценки риска необходимо дополнить ее вероятностными параметрами: частотой атак, зрелостью центра мониторинга безопасности, уровнем защищенности IdP, качеством журналирования, временем обнаружения и временем реагирования.

Кроме того, веса w_s, коэффициенты m_u и e_s требуют калибровки под конкретную организацию. Пример показывает возможный способ экспертного задания шкал, однако для промышленного применения модель необходимо настраивать по данным конкретной инфраструктуры: критичности сервисов, срокам действия маркеров, фактическим правам администраторов и числу доверяющих приложений.

Модель также не учитывает уязвимости кода конкретного IdP, ошибки реализации протоколов и внутренние угрозы со стороны легитимных администраторов. Эти аспекты должны рассматриваться в рамках отдельной модели угроз.

Переход от VPN к ZTNA снижает сетевую поверхность атаки, но не устраняет проблему концентрации риска. В ZTNA эта концентрация переносится из плоскости передачи данных в плоскость идентификации. Провайдер идентификационных данных

становится не просто сервисом аутентификации, а центральным элементом доверия, от которого зависит доступ к множеству приложений и сервисов. При этом настоящая работа не противопоставляет себя архитектуре ZTNA, а уточняет условия безопасного внедрения: снижение сетевой экспонированности должно сопровождаться защитой инфраструктуры идентификации.

Основные результаты работы состоят в следующем. Во-первых, выделена плоскость идентификации как самостоятельная составляющая плоскости управления ZTNA. Во-вторых, предложена сценарная модель, позволяющая отдельно оценивать компрометацию пользовательской сессии, ключа подписи, административного доступа и федеративного доверия. Показано, что наиболее опасные сценарии связаны не с захватом отдельной учетной записи, а с нарушением криптографического или административного доверия к IdP. В-третьих, введен коэффициент концентрации доверия ICF, позволяющий количественно сопоставлять различные сценарии по потенциальному охвату доверяющих сервисов.

Практическая значимость работы состоит в том, что предложенная модель позволяет обосновывать меры защиты IdP через уменьшение конкретных компонент компрометационного охвата: сокращение числа сервисов, доверяющих одному IdP; разделение административного и пользовательского контуров; защиту ключей подписи; ограничение федеративных связей; сокращение срока действия маркеров; применение устойчивой к фишингу многофакторной аутентификации.

Дальнейшие исследования целесообразно направить на эмпирическую калибровку коэффициентов модели по журналам IdP, средствам мониторинга и контроллерам ZTNA, а также на разработку методики автоматизированного расчета экспонированности IdP для корпоративных инфраструктур.

References

1. Syed N. F., Shah S. W., Shaghaghi A., Anwar A., Baig Z., Doss R. Zero Trust Architecture (ZTA): a comprehensive survey // IEEE Access. 2022. Vol. 10. P. 57143–57179. DOI: 10.1109/ACCESS.2022.3174679.

2. Zero Trust Architecture : NIST Special Publication 800-207 / Rose S., Borchert O., Mitchell S., Connelly S. National Institute of Standards and Technology, 2020. DOI: 10.6028/NIST.SP.800-207.

3. Assertions and protocols for the OASIS Security Assertion Markup Language (SAML) V2.0 : OASIS Standard / Cantor S., Kemp J., Philpott R., Maler E. OASIS, March 2005. URL: http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf (дата обращения: 07.05.2026).

4. OpenID Connect Core 1.0 incorporating errata set 2 / Sakimura N., Bradley J., Jones M., de Medeiros B., Mortimore C. OpenID Foundation, December 2023. URL: https://openid.net/specs/openid-connect-core-1_0.html (дата обращения: 10.05.2026).

5. Howard M., Pincus J., Wing J. M. Measuring relative attack surfaces / Lee D. T., Shieh S. P., Tygar J. D. (eds.) // Computer Security in the 21st Century. Boston : Springer, 2005. P. 109–137. DOI: 10.1007/0-387-24006-3_8.

6. Manadhata P. K., Wing J. M. An attack surface metric // IEEE Transactions on Software Engineering. 2011. Vol. 37, no. 3. P. 371–386.

7. Digital Identity Guidelines: Authentication and Authenticator Management : NIST Special Publication 800-63B-4 / Temoshok D., Fenton J. L., Choong Y.-Y., Lefkovitz N., Regenscheid A., Galluzzo R., Richer J. National Institute of Standards and Technology, 2025. DOI: 10.6028/NIST.SP.800-63B-4.

8. Multi-Factor Authentication Request Generation : Technique T1621 // MITRE ATT&CK. URL: https://attack.mitre.org/techniques/T1621/ (дата обращения: 16.05.2026).

9. Reiner S. Golden SAML: newly discovered attack technique forges authentication to cloud apps / CyberArk Labs, 2017. URL: https://www.cyberark.com/resources/threat-research-blog/golden-saml-newly-discovered-attack-technique-forges-authentication-to-cloud-apps (дата обращения: 16.05.2026).

10. Forge Web Credentials: SAML Tokens : Technique T1606.002 // MITRE ATT&CK. URL: https://attack.mitre.org/techniques/T1606/002/ (дата обращения: 17.05.2026).

11. Adversary-in-the-Middle : Technique T1557 // MITRE ATT&CK. URL: https://attack.mitre.org/techniques/T1557/ (дата обращения: 20.05.2026).

12. Microsoft mitigates China-based threat actor Storm-0558 targeting of customer email / Microsoft Security Response Center. Microsoft, 2023. URL: https://www.microsoft.com/en-us/msrc/blog/2023/07/microsoft-mitigates-china-based-threat-actor-storm-0558-targeting-of-customer-email (дата обращения: 21.05.2026).

13. Implementing Phishing-Resistant MFA / CISA. Cybersecurity and Infrastructure Security Agency, 2022. URL: https://www.cisa.gov/sites/default/files/publications/fact-sheet-implementing-phishing-resistant-mfa-508c.pdf (дата обращения: 25.05.2026).