PYTH 网络：

第一方金融市场数据预言机

Pyth 数据协会

1.0 版本

2022 年 1 月 4 日

金融市场数据通常仅供特定的机构和用户访问。传统市场通常对实时喂价（*price feeds*）和历史价格信息保持着严格的控制。尽管加密货币市场目前数据壁垒较少，但无法保证这种状况会持续下去。最终，只有特定用户群体能获取最及时、最准确、最有价值的信息。

Pyth 网络旨在将这些宝贵的金融市场数据带到去中心化金融（*DeFi*）应用和公众视野中。为了实现这一目标，Pyth 为金融交易机构、做市商、交易所等市场参与者提供经济回报，激励它们分享在运营中收集的价格数据信息。Pyth 网络聚合一手的价格数据并在链上发布，供链上或链下应用使用。

面向这一目标，Pyth 网络已经取得了实质性进展。我们的初始版本正在 Solana 链上运行，以亚秒级速度更新美国股票、外汇交易、大宗商品和加密货币的价格数据。Pyth 网络的数据提供商包括传统金融和加密货币领域的许多知名交易公司和交易所。我们已经为 Solana 链上的一些头部协议提供了数据喂价，并将很快集成其他区块链网络。

本白皮书旨在详细阐述 Pyth 网络背后的协议设计，其目的在于使 Pyth 网络实现自我维持和去中心化。Pyth 的协议包含一系列机制和激励措施，用于网络参与者进行统筹协调[1]。本白皮书将对 Pyth 网络中各类参与者的角色以及相关协调机制进行说明。

[1] 在本白皮书中，“网络”一词指 Pyth 协议及其参与者的具体实例。当“Pyth 网络”执行某种行为时，该行为实际上是由参与者通过与协议实例进行交互来执行的。

本白皮书由 Pyth 数据协会（简称“协会”）与 Pyth 网络的参与者社群共同发布，描绘了 Pyth 网络的未来愿景。协会和网络参与者将根据本白皮书中的理念、更广泛的加密社群的反馈以及 PYTH 代币持有者的治理意见，为协议的初步发展提供指导。

1 概述

Pyth 协议旨在激励参与者持续发布各种货币对（如 BTC/USD）的最新价格数据。每种产品均有喂价（当前价格不断更新）和置信区间（表示价格预测的不确定性）。例如，当前 BTC/USD 的喂价表明其价格为 65000 ± 50 美元。每种产品的喂价都将发布在链上，数据消费者（consumer），即区块链程序或链下应用，可以读取喂价信息。链上程序整合各个数据发布者（publisher）的喂价，生成每种产品的聚合喂价。协议旨在吸引第一方数据的所有者成为数据发布者，发布者须具备及时获取高质量价格信息的能力。

在发布喂价之上，Pyth 协议允许消费者选择性地支付数据费用（data fee）。作为支付数据费用的回报，如果预言机发布的价格不准确，消费者可以从委托者那里获得补偿款。数据费用帮助那些使用 Pyth 喂价的项目对冲预言机的不准确性风险。部分数据费用将支付给数据发布者，作为发布数据的回报。

因此，Pyth 协议具有以下三类参与者：

- **发布者（Publisher）**负责发布喂价，并获得部分数据费用作为回报。发布者通常是能够及时获取准确价格信息的市场参与者。Pyth 协议根据发布者分享的新价格信息的数量，按比例对发布者进行奖励。
- **消费者（Consumer）**读取喂价，将数据集成智能合约或去中心化应用之中，并可以选择性地支付数据费用。消费者既可以是链上协议，也可以是链下应用。
- **委托者（Delegator）**质押代币，赚取数据费用，代价则是在预言机价格不准确的情况下，可能会损失其质押的代币。

注意，单一主体可能会在 Pyth 协议中扮演多个角色；例如，发布者可能同时也是委托者。

这些参与者通过四种机制进行互动。所有机制都会在链上实施：

- **价格聚合机制（Price aggregation）**将单个发布者的喂价整合为产品的单一喂价。此机制旨在生成稳定的喂价，也就是说，喂价不会受到少数发布者的显著影响。
- **数据质押机制（Data staking）**使委托者能够通过质押代币来获取数据费用。委托者总体上还通过质押代币的方式决定每个发布者对聚合价格的影响程度。另外，此机制还将确定委托者质押的代币是否会遭到削减。最后，此机制向消费者收取数据费用，并向委托者分配部分数据费用。余下部分将进入奖励池，用于分配给发布者。
- **奖励分配机制（Reward distribution）**决定了每个数据发布者可从奖励池中获取的奖励份额。此机制会优先奖励提供高质量喂价的发布者，并降低提供低质量信息的发布者获取奖励的可能性。
- **治理机制（Governance）**决定上述三种机制的高级参数。

如何确保上述机制抵御各种攻击行为的稳健性，是 Pyth 面临的关键挑战。以下三类攻击行为需要重点防范：

1. 参与者可能会以发布者的身份加入，试图操纵预言机的价格数据。价格聚合机制旨在通过限制单个发布者对聚合价格的影响来防范这种攻击。
2. 提供低质量信息的参与者可能会以发布者的身份加入，不提供有用的价格信息，平白获取奖励。奖励分配机制旨在通过降低提供低质量信息的发布者获取奖励的可能性来防范这种攻击。
3. 参与者可能会先支付数据费用，而后试图通过操纵索赔流程来触发赔偿。这些机制将设计恰当的索赔流程，让这种攻击变得难以实施。

上述机制依赖 Pyth 协议的两大核心功能。首先，Pyth 协议的部分内容将以纪元块（*epoch*）的方式运行。一个纪元块包含一系列 Solana 区块时段（*slot*），每个纪元块相当于现实世界的一周。其次，Pyth 协议将要求用户质押 PYTH 代币才能参与某些活动。用户的代币在质押时会立即锁定，并在下一个纪元块开始时用于下游活动。质押者可以随时要求解除质押。质押解除后，代币将在当前纪元块和下一个纪元块内

锁定在合约中。这种质押设计可以保证为任何特定活动质押的 PYTH 代币数量在一个纪元块内保持不变。协议另有其他机制，例如质押池机制允许质押者将其质押代币委托给其他用户，但这些机制并非协议的核心部分（并可作为完全独立的程序进行搭建），因此，本白皮书没有对这些机制进行说明。

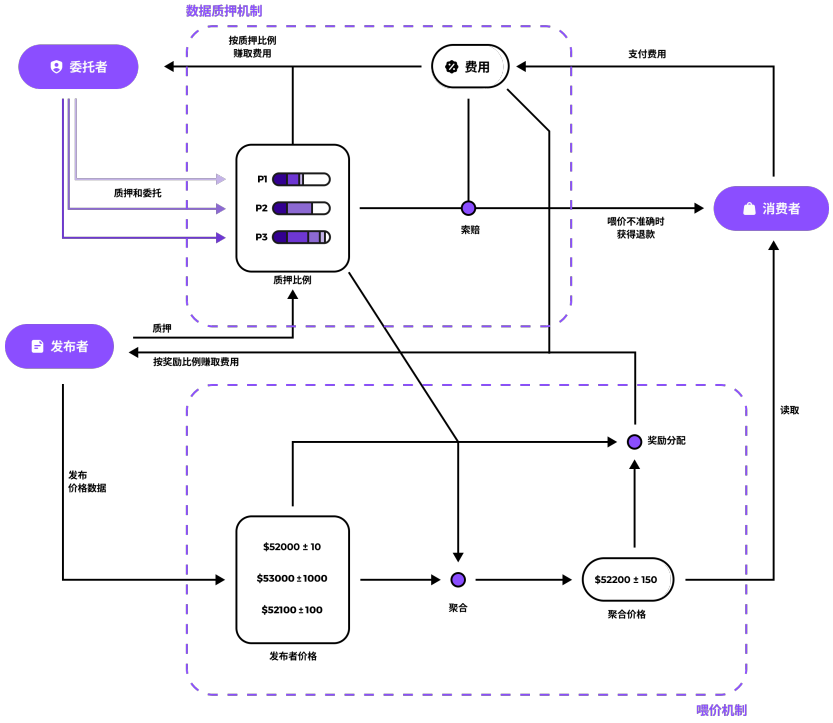

图 1： Pyth 协议概览。图 1 呈现了参与者（紫色椭圆形）及其与各种机制（紫色虚线框）之间的相互作用。各机制的详细信息请参阅正文。

2 价格聚合

价格聚合机制将每个发布者的价格和置信区间组合成单一的聚合价格和置信区间。例如，一个发布者说 BTC/USD 的价格是 52000 ± 10 美元，另一个发布者说 BTC/USD 的价格是 53000 ± 20 美元，价格聚合机制可以将这两个价格聚合成 52500 ± 500 美元的聚合价格。此机制是链上程序的组成部分之一，在在发布者提交

价格更新时触发：给定区块时段的首次价格更新，会自动整合前一区块时段的聚合价格。

价格聚合算法具有以下三种特性：

1. 不受发布者的蓄意或无意操纵行为的影响。例如，如果大多数发布者提交的价格为 100 美元，而一个发布者提交的价格为 80 美元，则聚合价格仍接近 100 美元。即使有一小部分发布者提交的价格与行情相去甚远，这种特性也能增加聚合价格保持准确的可能性。图 2（a）即呈现了这种情形。
2. 聚合价格对准确性不同的数据源进行适当地加权。由于发布者观察到的产品价格准确性不同，Pyth 协议允许发布者提交其对发布的产品价格的置信程度。例如，一些发布者本身是交易所。而交易所之间流动性水平各异，相较于流动性较高的交易所，流动性较低的交易所往往买卖价差（bid/offer spreads）更大。因此，可能会出现一个交易所报价 101 ± 1 美元，而另一个交易所报价 110 ± 10 美元的情况。经过加权后，聚合价格将会更接近 101 美元而并非 110 美元。图 2（b）即呈现了这种情形。
3. 聚合置信区间反映了发布者之间的价格差异。事实上，任何产品均不存在单一价格。在给定的任何时间，一个产品在不同场所的交易价格会略有不同。这种价差也是对产品价格准确性的根本限制。聚合置信区间可以反映不同场所间的差异和这些限制因素。图 2（c）和（d）即呈现了交易所之间存在价格差异的两种情况。

价格聚合算法采用的是加权中位数的一种变体。算法的输入值是每个发布者的质押权重（stake-weight）。该权重由数据质押机制（详见下文）生成，用于最大限度地提高喂价的稳健程度。权重用于反映可能影响稳健性但难以量化的因素（例如发布者的声誉）。算法的第一步是通过给每个发布者三张表决票来计算聚合价格，其中一张用于对其自己的价格进行表决，另外两张用于对其价格 ± 置信区间后的数值进行表决，然后取表决结果的质押加权中位数。第二步是计算从聚合价格分别到表决结果质押加权后数值的第 25 个百分位和第 75 个百分位所对应的值的距离，然后选择其中较大者作为聚合置信区间。

这一简单算法是对普通中位数进行归纳处理。大多数人将中位数理解为数据集的中间值，即第 50 个百分位所对应的值。然而，中位数也是使目标函数 $\sum_i |R - p_i|$ 出现最小值时所取的 R 值，其中 p_i 是第 i 个发布者的价格。该函数根据 R 与发布者价格 p_i 之间的距离对 R 进行惩罚。这一算法用于计算使 $\frac{1}{3}\sum_i s_i|R - p_i| + \frac{2}{3}\sum_i s_i \max(|R - p_i| - c_i, 0)$ 出现最小值时所取的聚合价格 R，其中 s_i 为发布者的质押权重、c_i 为发布者的置信区间。这样做有两个目的。首先，根据发布者的质押额度对发布者进行加权，这样质押额度低的发布者对价格的影响最小。其次，将普通中位数目标与函数的第二项相结合，第二项仅在 R 落在置信区间外时对其进行惩罚。

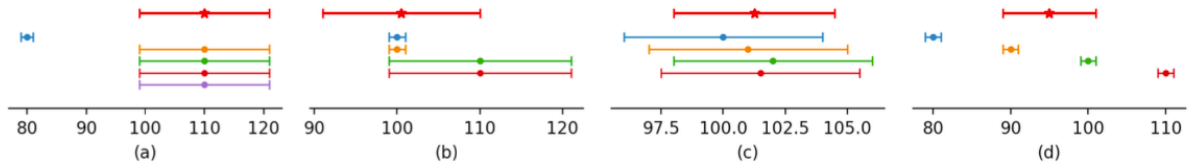

图 2: 聚合过程的情形。图中下方较细线条表示每个发布者的价格和置信区间，上方较粗红色线条表示最终的聚合价格和聚合置信区间。

3 数据质押

数据质押机制负责收取数据费用，如果预言机发布的产品价格不正确，则将分配退款[2]。此机制还将定义索赔流程，用于确定消费者何时收到数据质押退款；索赔流程将预言机的价格与外部参考数据进行比较，以衡量预言机发布价格的准确性。

数据质押机制的设计面临着以下挑战。首先，数据费用需要合理定价，即价格应反映数据错误的风险。其次，此机制需要协调发布者和委托者的激励机制，因为发布者出现错误可能会导致委托者遭受损失。最后，索赔流程应当具备稳健性，以防止虚假索赔行为。

[2] 需要注意的是，本白皮书中不考虑宕机的情形，即预言机无法发布价格的情况。本白皮书未来可能会涵盖由发布者错误和非底层区块链故障导致的宕机情形。

数据质押机制将允许或要求用户执行以下行为：

1. 消费者可以选择为产品支付数据费用。消费者可以使用经过治理机制批准的任何代币向 PYTH 协议支付费用，可能包括 PYTH、USDC 或其他代币。作为交换，如果消费者选择的产品在接下来的 4 个纪元块（约 1 个月）内出现问题，消费者可以获取退款。
2. 发布者需为其喂价的每种产品质押不低于特定数量的 PYTH 代币。这些 PYTH 代币是退款的来源之一，但只有在发布者发布的价格不准确时才会削减。换言之，只有当链上机制确定某个发布者出现错误时，协议才会对该发布者进行惩罚。
3. 委托者可以使用质押中的PYTH代币来为产品担保。委托者可为质押中的每个PYTH代币选择（1）支持某产品（2）支持该产品的某个发布者。选择支持发布者有助于提升产品的安全性（详见下文）。委托者从他们提供支持的产品中获取部分数据费用；治理机制将决定相应的份额（初始份额为 80%）。
4. 任何人都可以就预言机发布的错误价格提出索赔（claim）。一旦有人提出索赔，PYTH 代币持有者将通过表决投票来决定索赔是否生效。如果索赔成功，协议将会削减产品委托者质押的代币。协议将按照索赔事件所属纪元块内消费者支付的数据费用（按美元计算），将削减的金额按比例分配给消费者。如果在任何期间内 PYTH 的聚合价格和聚合置信区间均存在显著错误，索赔裁定流程会判定索赔成功。该流程将进一步确定在此期间发生错误的发布者，并相应削减他们的质押代币。

此功能将作为链上程序的一部分进行实施，成为 Pyth 协议的规则之一。例如，消费者会调用链上程序中的函数来对冲产品错误风险。此功能会接受付款并存储消费者的钱包地址，以便未来进行退款。索赔流程涉及两种独立的链上行为：提出索赔和批准索赔。其中，提出索赔需要通过链下计算得到特定输入数据；Pyth 网络的开发者将会提供开源软件包来构建这些数据。索赔流程的激励措施旨在确保这些链下活动得到正确执行。

数据质押机制旨在为数据费用制定一个合适的市场价格。在步骤（3）中，委托者会比较预期数据费用与成功索赔的预估风险。高费用低风险会吸引更多的委托者选择该产品，从而降低数据费用。随着这一流

程达到平衡，数据费用与其总质押数额之间的关系应能反映其索赔风险。

委托者在此过程中还起着另一个作用，即负责确定发布者在聚合过程中的质押权重。这一流程也发生在步骤（3），即委托者选择其支持的产品发布者。一个发布者的质押权重是该发布者的质押额与其支持者的质押额之和，归一化后每个产品的总质押权重为 1。但需要注意的是，委托者是为产品的错误担保，不是为特定的发布者担保，也就是说，如果其他发布者出现错误，协议仍将在有关索赔中削减委托者的质押代币。因此，协议鼓励委托者支持多个发布者，从而最大程度降低产品的整体错误风险。委托者需要平衡多种相互冲突的因素来作出决定。一方面，历史表现和声誉良好的发布者值得更大的质押额；另一方面，将质押额分配给多个发布者可以确保少数发布者出现错误不会导致整个产品的数据错误。

治理机制可能会允许委托者用每个质押中的 PYTH 代币为多个产品提供支持。这将会使数据质押更具资本效率，因为多个产品同时遭到索赔是极小概率事件。

3.1 索赔

索赔流程将决定是否发生退款。索赔流程的目的在于确认产品的聚合价格和聚合置信区间与一些真实链下数据相比是否有误。索赔流程很难设计，因为索赔成功可能会对 PYTH 代币持有者产生不利影响。因此，这一流程中 PYTH 代币持有者不会是公正的鉴定者。另一方面，攻击者也可能会收买公正的鉴定者，即完全没有利害关系的主体，来操纵错误的退款。

下述索赔流程将缓和上述两个矛盾问题。此流程将使用 HUMAN 协议向公正的鉴定者收集必要的链下信息，然后将这些信息输入到预设算法，确定索赔的结果。PYTH 代币持有者随后将通过表决对其进行批准。批准后，即进行退款。这一机制借助社会压力来激励 PYTH 代币持有者出于协议的长期利益行事。简单地说，如果 PYTH 代币持有者不予批准退款，算法的结果将有力地向消费者证明数据费用是没有用的。结果就是没有人会继续支付数据费用，整个协议都可能会失效。但如果攻击者试图收买鉴定者，PYTH 代币持有者可以通过表决反对作出批准，并公开攻击者的操纵证据行为。

任何人都可以就协议提出索赔，并（可能）获得退款。提出索赔即主张聚合价格和聚合置信区间在特定时间段存在错误。为防止恶意索赔，索赔者需提供一些 PYTH 代币作为保证金；如果索赔获得批准，协议将会退还保证金。索赔者还需预付 HUMAN 任务的有关费用；如果索赔获得批准，协议将会偿还这笔预付款项。

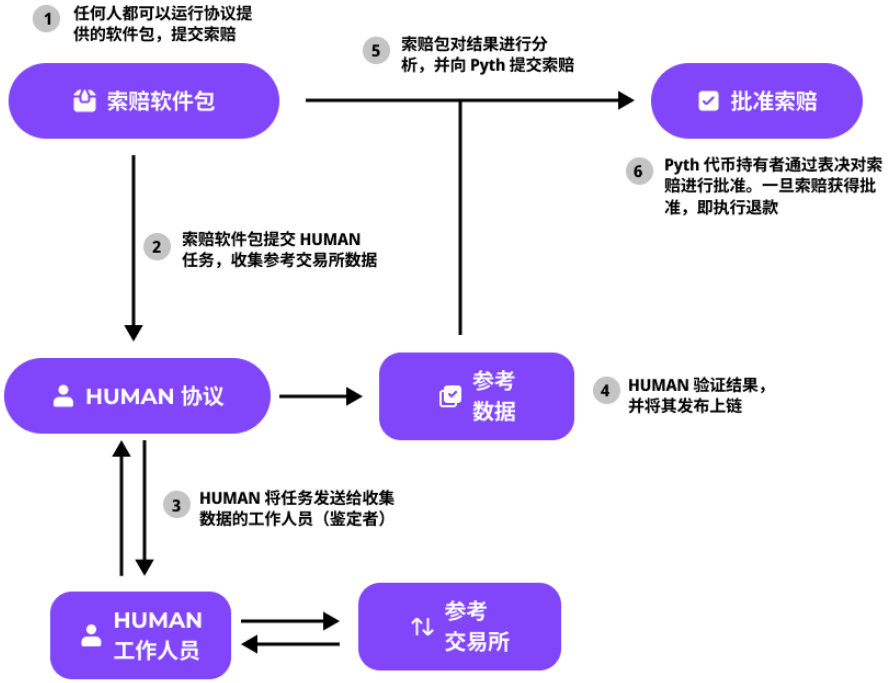

图 3：索赔流程步骤图。

索赔应包括以下字段：

1. 出现数据错误的产品
2. 出现数据错误的时间间隔。时间间隔应相对较短，例如 1 秒钟。

3. HUMAN 协议任务就索赔的真实性及发布者有无错误得出结果。Pyth 网络的开发者将提供一个 GitHub 数据库，任何人都能够快速运行必要的 HUMAN 任务，并以可验证的方式组合其结果。

HUMAN 任务将随机抽取工作人员，对下列链下信息进行报告：

1. 由一组固定的参考交易所给出的产品在相关时间间隔内的最高价格和最低价格。Pyth 协议的治理机制将为每个产品提前选定参考交易所。
2. 相关时间区间内的最大聚合价格和最小聚合价格以及最大置信区间。
3. 相关时间区间内每个发布者的最高价格和最低价格以及最大置信区间。

符合下列条件时，索赔将获得成功：（1）喂价在索赔对应的时间区间内发布了一个聚合价格，并且（2）所发布的价格（将置信区间纳入考虑的前提下）与参考价格不一致。计算一致与否的算法会比较两个价格范围。首先是 Pyth 网络的价格范围，其下限为最低聚合价格减去 3 倍置信区间，上限为最高聚合价格加上 3 倍置信区间。其次是参考价格范围，其下限为最低参考价格，上限为最高参考价格。如果这两个范围互不重叠，算法将判定索赔成功，因为这种情况下 PYTH 网络几乎不可能给出符合参考价格的喂价。如果索赔成功，算法会另外识别存在错误的发布者。算法将用相同的标准（判断范围是否重叠）来对发布者各自的报价和置信区间进行检验。其计算结果将（1）对索赔是否成功作出是/否的结论，及（2）如果得出肯定结论，则将确定存在错误的发布者。

HUMAN 任务的配置将提高攻击者影响索赔流程的难度。具体而言，任务会立即分配给全球各地的鉴定者，鉴定者需在一定程度上达成一致后方可得出最终答案。HUMAN 协议将签署该任务的结果并在链上发布，PYTH 代币持有者可以在批准步骤中对这些结果进行验证。

索赔流程的这一部分将包含在软件包中，任何人都可以运行。软件包将会（1）利用适当的参数将必要的 HUMAN 任务实例化，（2）向 HUMAN 协议的用户收集结果，及（3）运行上述算法以确定索赔是否成功。软件包用户可以直接提交其输出值，进入批准流程。

最后，PYTH 代币持有者将通过表决对算法的输出值进行批准。代币持有者应对 HUMAN 任务的链上结果进行确认，然后运行自己的索赔软件实例，确认是否相符。代币持有者不应姑息任何试图操纵索赔流程的情形，例如，如发现收买行为，应通报给 HUMAN 鉴定者（HUMAN 任务的配置决定了代币持有者往往是能够发现收买行为的存在的）。如果索赔结果属实，代币持有者应通过表决批准索赔。批准后，协议将会削减产品委托者和过错发布者的质押代币，并把削减的代币分配给支付数据费用的消费者。

从 PYTH 协议链上执行的角度来看，索赔流程仅分为两个步骤。第一步是提交索赔，第二步是批准索赔。这两个步骤都不必然地需要链下计算或讨论，但协议会鼓励各方执行这些链下活动。例如，用户可以选择不运行 HUMAN 任务，直接提交索赔，但 PYTH 代币持有者不太可能对该索赔作出批准。同样，PYTH 代币持有者可能会通过表决来否决每一项索赔。然而，当 HUMAN 任务的结果表明索赔应该成立时，对 PYTH 代币持有者而言，批准索赔会有利很多。

4 奖励分配

奖励分配机制将确定每个发布者获得的奖励池份额。如前文所述，数据质押机制会将每款产品的部分数据费用分配到该产品的奖励池中。这个比例最初将设置为 20%，可以通过治理机制进行调整。奖励池还可能包括用于促进协议发展的额外奖励（详见下文）。

奖励分配机制旨在实现以下目标：

1. **优先奖励质量更高的发布者**。发布者质量各异，有些发布者能获得比其他发布者更准确或及时的价格信息。奖励系统应优先奖励这些发布者，以便激励最佳发布者为协议做出贡献。
2. **防止攻击者获得奖励**。发布者可能会试图利用这一系统谋取私利。奖励机制必须惩罚攻击者，防止他们参与 Pyth 协议。

3. **鼓励诚实报告非公开信息**。发布者通常拥有非公开的价格信息，例如他们在各交易所的最新交易。激励措施应鼓励发布者使用这些非公开信息诚实地报告价格，因为发布者的诚实报告能帮助 Pyth 生成最准确的聚合价格。

现有的预言机机制无法实现这三个目标。其他预言机会奖励报告数据与其他发布者一致的发布者，即报告相同价格的发布者。然而，激励机制以一致性为基础，意味着发布者可以通过谎报非公开价格信息牟取不当利益。例如，某个发布者认为当前价格为 110 美元，但它注意到前一个区块时段的聚合价格为 100 美元。该发布者知道价格不太可能在单个区块时段中变动 10 美元，因此，它可以推断当前区块时段的聚合价格也极有可能在 100 美元左右。于是，发布者可能会选择报告接近 100 美元的价格来最大化其价格与聚合价格的一致性，而非其实际估计的 110 美元。由此可见，激励机制以一致性为基础，将促使发布者报告*它们对其他发布者的报价的最佳估计*，而不是它们自己掌握的非公开价格。

第二个问题在于，激励机制以一致性为基础，将无法分辨恶意参与者和善意参与者。价格通常会在一段时间内保持稳定，所以上一个 Solana 区块时段的聚合价格是对当前价格的合理估计。因此，如果以一致性为发放奖励的基础，恶意参与者便可通过“重新播报”聚合价格来轻松获得奖励。预言机协议也不应简单地惩罚报告价格不一致的发布者，因为诚实的发布者偶尔会偏离实际价格。

之所以会出现这些问题，是因为现有的预言机是针对每个发布者都能访问相同信息的情况而设计的。在这些情况下，以一致性为发放奖励的基础足以证明发布者已正确报告信息，因为诚实的发布者没有理由出现不一致的情况。但在 Pyth，发布者拥有非公开信息，报告的价格预计不会完全相同。现实中也存在各交易所价格不一致的情况，预言机应该反映这些情况。

本白皮书根据批判性见解，针对非公开信息设置提出了一种新的预言机机制：发布者应当因分享新的信息（即更新当前价格）而获得奖励。这个机制通过计算价格序列对未来聚合价格变化的预测准确程度来衡量新信息的价值。恶意参与者很难通过这种机制攫取利益，因为它们无法轻易地根据历史价格预测未来价格。

4.1 分数计算

奖励分配机制将按以下三个数值的比例，将产品的奖励池分配给发布者：

1. 发布者的质押权重 s，该数值由数据质押机制确定，介于 0 和 1 之间。
2. 质量得分 q，该数值衡量发布者的价格序列对于未来价格变化的预测准确程度，介于-1 和 1 之间。正分数代表着一定的预测能力。
3. 发布者置信区间的校准值 c。该数值介于 0 和 1 之间。

奖励分配机制将在每个纪元块结束时分配奖励。链上程序将评估每个发布者在该纪元块期间的 q 和 c 值。在纪元块结束时，每个发布者将按照 $s \times q \times c$ 的比例分得奖励池的一部分。如果这个值是负数，奖励分配机制将削减发布者相应数量的质押代币。

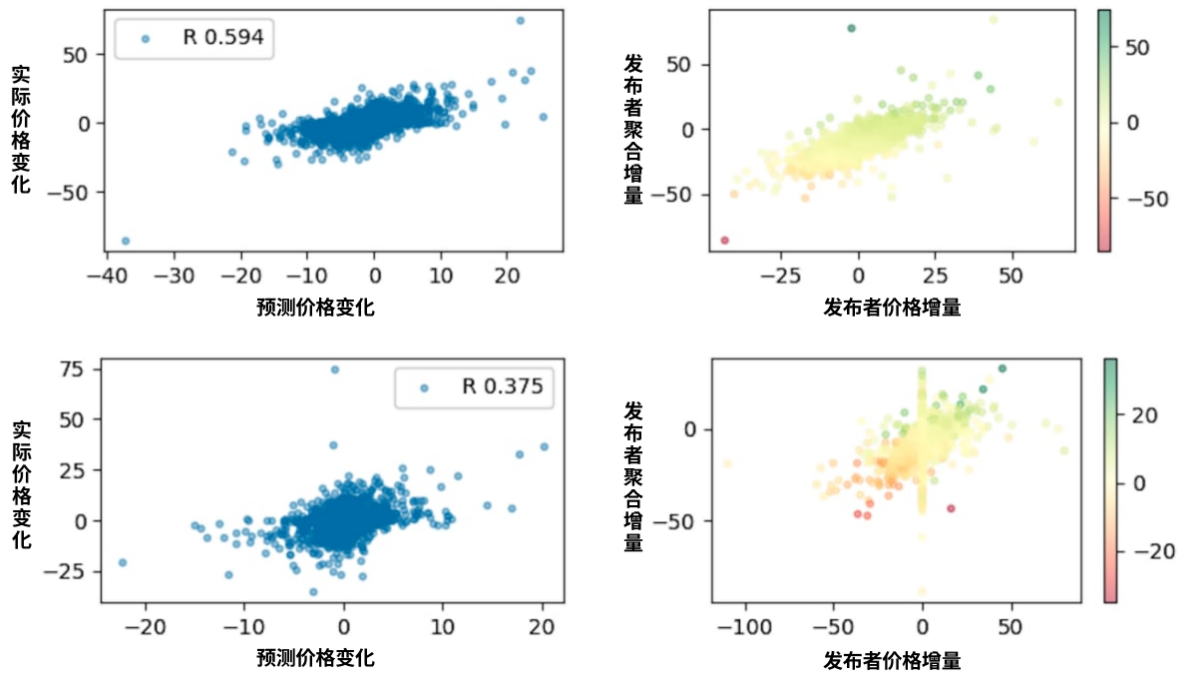

图 4: 历史数据样本中两个发布者的质量分数。左图是预测和实际价格变化的散点图。两个数值之间的线性关系越紧密，质量分数就越高。右图的横轴和纵轴对应回归模型的两个无偏差特征，颜色表示实际价格变化。颜色渐变平滑表示高质量分数。在此示例中，上方的发布者比下方发布者的分数更高。

4.1.1 质量分数

质量分数衡量一个发布者的价格序列对聚合价格未来变化的预测准确程度。奖励分配机制将训练一个在线回归模型来计算这一数值，该模型会根据发布者价格序列的几个特征来预测未来价格。设 p_t, σ_t 为发布者在区块时段 t 的价格和置信度，设 $\bar{p_t}$ 为该区块时段的聚合价格。也就是说，聚合算法根据所有发布者的价格 p_t 计算 $\bar{p_t}$。在区块时段 t，回归模型执行以下更新：

$$
\begin{aligned}
f_t &\leftarrow [p_t - \bar{p}_{t-1}, p_t - p_{t-1}, 1] \\
\hat{p_t} &\leftarrow \omega_{t-1}^T f_t + \bar{p}_{t-1} \\
\omega_t &\leftarrow clip\left(\frac{\alpha}{\sigma_t}(\hat{p_t} - \bar{p_t}) f_t, -0.1, 0.1\right)
\end{aligned}
$$

在上述等式中，$\hat{p_t}$ 表示对当前时间步内的聚合价格预测。线性回归模型根据两个价格变化特征计算此预测值，这些特征比较发布者的当前价格 p_t 与其自己的价格，以及前一个区块时段的聚合价格。ω_t 表示回归模型的权重。每次预测之后，最终方程会使用标准的裁剪梯度更新这些权重。

质量分数 q 是两项的乘积：第一项是预测价格 $\hat{p_t}$ 与整个纪元块内聚合价格 $\bar{p_t}$ 之间的相关性，第二项是发布者提供价格的区块时段比例。相关性衡量预测价格与聚合价格的符合程度。如果预测是完美的，则相关性为 1。如果预测是随机的，则相关性为 0。如果（不知何故）预测比随机更差，则相关性为负。诚实的发布者不太可能遇到第三种情况，这种设计用于防止攻击行为（详见下文 4.2 节）。

这里所说的质量分数可以有所变化。首先，它可能会预测超过下 1 个区块时段的价格变化，而这应该是更加难以预测的。其次，它可以包含额外的功能。例如，一些交易所通常存在微小但持续的价格差异。如果这一机制包含追踪价格差异的功能，则可以避免对此类价格差异进行惩罚。我们的开发人员已经在历史数据上实验了这种机制。实验表明，本白皮书所所述的版本产生了合理的结果。当然，方法是灵活的，将来可以再做调整。

4.1.2 校准分数

校准分数 c 衡量发布者的置信区间能否准确反映其数据的不确定性。校准分数将置信区间解释为发布者期望在其中找到聚合价格的拉普拉斯分布（*Laplace distribution*）标准差（拉普拉斯分布是一种重尾分布，相比于正态分布，它能更好地反映价格的实际分布情况）。校准分数使用简单的频率测试来衡量聚合价格与理论上的拉普拉斯分布的接近程度。具体来说，它利用聚合价格和预测值之间的差异计算 z 分数，再通过发布者的置信区间进行归一化。得到的 z 分数随后会归入统计堆，生成直方图；每个统计堆的 z 分数阈值都已选定，以使每个统计堆在标准拉普拉斯分布下具有相等的概率。对于完美的发布者，这一过程应该生成一个均匀分布的直方图。因此，校准分数被定义为 1 减去发布者的直方图到均匀分布的直方图之间的 EMD 距离。

请注意，校准分数并不代表发布者能够通过提供更窄的置信区间获益。质量分数已经解决了这个问题：置信区间更窄的发布者的价格预测应该更加准确。相反，校准分数 c 表示的是发布者所报告的置信区间是否与其“真实”置信度一致。

4.2 讨论

这种奖励分配机制有助于上文所述目标的实现：

1. 此机制为价格序列预测性更高、置信区间校准度更好的发布者分配更高的奖励。这些发布者也更有机会得到更高比例的质押权重。于是，这些发布者将分得奖励池的更大份额。由此，奖励分配机制可激励最佳发布者参与协议。
2. 相关系数的特性降低了恶意参与者在系统中获得奖励的可能性。这一点建立在两个假设之上：（1）未来价格很难通过历史价格预测，并且（2）Pyth 聚合价格将追踪实际价格。第一个假设属于有效市场假说，它在所需的小时间尺度上大体是正确的。第二个假设也是成立的，因为聚合算法具备稳健性。因此，攻击者操纵 Pyth 聚合价格偏离实际价格的能力有限。在这两个假设下，只能访问 Pyth 历史数据的攻击者无法作为发布者获得奖励：它们的预测 $\hat{p_t}$ 是随机的，其相关性 q 也将是一个零期望的随机变量。每个发布者的奖励/惩罚与 q 成正比，因此，它们在每个纪元块中的预期奖励均

为零。此外，奖励机制通过削减质押数额来惩罚质量得分为负的发布者，因此，它们在多个纪元块中的预期奖励都是零。惩罚机制是必不可少的：否则恶意人士可能会在多个纪元块中进行恶意操作，通过投机在部分纪元块中获得正的质量分数。惩罚机制将让它们其他纪元块中遭到削减，抵消依靠投机获得的奖励。

3. 预测值的适应性消除了不诚实行为的主要诱因。只要回归模型计算的预测值与聚合价格一致，发布者就不必与价格聚合价格保持一致。

这种奖励机制确实有一些小的弱点。第一个弱点是攻击者可在同一区块时段复制发布者的价格。这种攻击需要攻击者读取发布者的价格更新，并在同一 Solana 区块时段内提交其自己的价格更新。Pyth 协议可提交-显示机制（*commit-reveal*）来防御这种攻击。第二个弱点是此机制鼓励发布者提交对未来价格的预测，而这可能与实际价格不一致。这种攻击似乎不太可能发生：如果发布者可以预测未来的价格，直接利用这些信息开展交易的收益比发布数据高得多，也不会面临被数据质押机制削减质押的风险。

5 治理

Pyth 数据协会将在协议处于开发阶段时对其进行初期管理。随着时间的推移，协会会将协议的完整控制权转移给链上的治理机制。转移后，所有协议治理都将在链上进行。然而，链上治理系统存在各种缺陷，例如借币投票——这目前还没有有效的技术解决方案。因此，协议的总体设计理念是减少治理输入的必要性。

链上治理机制将使用代币投票系统批准或拒绝提案。任何达到 PYTH 代币质押门槛的人都可以提出治理提案。协议将让 PYTH 代币质押者对这些提案进行投票。代币质押者还可以将自己的选票委托给其他人。治理投票将从提案之日起持续 2 周，并且仅允许在之前的纪元块中质押代币的持有者进行投票。用户可以使用为其他用途质押的 PYTH 代币进行投票（例如数据质押）。此外，部分用户的 PYTH 代币处于锁定状态，这些代币也可用于投票。

以下行动预计均将通过链上治理执行：

- 批准可用于数据费用的代币类型。

- 确定在 PYTH 上线的产品及其参考数据（例如，价格保留几位小数、参考汇率）。
- 确定分配给发布者、委托者和其他用途的数据费用份额。
- 批准链上程序的软件更新。
- 确定发布者质押 PYTH 代币的数量门槛。
- 确定委托者可用每个质押中的 PYTH 代币支持多少产品。
- 允许针对产品提出索赔。一旦产品有足够的发布者来产生稳健喂价，治理机制就应采取这一行动。
- 允许发布者提供喂价。

6 激励措施

本节总结了协议中各种参与者的激励措施。

协议通过数据质押和奖励分配机制，激励**发布者**发布及时准确的价格数据。参与者必须质押 PYTH 代币才能成为发布者。如果（1）其发布的价格与参考价格相差甚远，且（2）该价格随后导致预言机的喂价出现问题，则数据质押机制可能会削减其质押的代币。这种可能的惩罚会鼓励发布者发布接近整体市场的价格。奖励分配机制还会优先奖励其发布价格预测了聚合价格的未来变化的发布者。

协议通过将部分奖励分配给发布者，激励发布者的参与。发布者分得其发布价格的产品的部分数据费用作为收益。产品的数据费用很可能会与消费者对该价格信息的使用成比例增长，而发布者投入的资本（质押的 PYTH 代币）和发布成本保持不变。发布者数量少但价格数据用量高的产品会产生很有吸引力的数据费用。

发布者可能执行的主要攻击包括（1）试图操纵聚合价格，及（2）试图在不提供新定价信息的情况下获得奖励。为防止第一类攻击，委托者将设置每个发布者的质押权重，限制它们对聚合价格的影响。协议激励委托者设置这些权重，由此，任何个别发布者或小团体都无法操纵价格。为防止第二类攻击，奖励机制仅奖励其价格序列预测了未来价格变化的发布者。

协议通过以下两点激励消费者支付数据费用。首先，数据费用能降低应用程序使用 PYTH 喂价的风险。应用程序的用户对风险高度敏感，并会相应地做出使用决策。其次，支付数据费用可以吸引更多的发布者提供该产品的价格信息，从而提高喂价的稳健性。

消费者可能会想先支付数据费用，再攻击系统，申请赔偿。但索赔流程经过了精心设计，能降低出现这种结果的可能性。

协议通过数据费用激励委托者参与协议。委托者初期将获得有吸引力的报酬，但随着市场效率提高，委托者之间的竞争将越发激烈，报酬将相应减少。从长远来看，委托者获得的报酬将反映委托活动的固有风险和更广泛的投资环境。

委托者还有另一个作用，即设置发布者的质押权重。协议激励委托者以最大化喂价稳健性为目标来设置这些权重，因为如果聚合喂价不正确，委托人可能会失去其质押的代币。

7 代币分配

PYTH 代币的总供应量为 10,000,000,000，不会增加。85%的代币最初会在合约上锁定。这些代币有 1 年的锁定期，将在 7 年内每月线性解锁。这将随着时间的推移逐渐增加非锁定代币的供应量。余下的 15%PYTH 代币最初就是非锁定的状态。锁定/非锁定代币将按照表 1 所列的门类进行分配。

	非锁定	锁定	总计
链上奖励	8%	14%	22%
生态参与	5%	28%	33%
团队和顾问		25%	25%
启动伙伴	2%	8%	10%
私募销售		10%	10%

表 1：锁定/非锁定 PYTH 代币的分配。锁定代币有 1 年的锁定期，并将在 7 年内每月线性解锁。

7.1 启动奖励

Pyth 协议很可能会为早期参与者提供额外的激励。具体来说，协议面临新产品的冷启动问题。这些产品既没有发布者发布数据，也没有消费者支付数据费用。没有数据费用，发布者也缺乏发布数据的动力。为新产品的早期发布者发放奖励，是协议此问题的方法之一。

协议可以采取多种激励系统来解决冷启动问题。例如，协议可以投入额外的代币到新产品的奖励池。这些代币可以来自预先分配的奖励池或成熟产品的数据费用。协议可以采用多种方式来分配这些额外的代币，例如，根据获取不同类型数据的相对难度进行分配。协会将保留 PYTH 代币供应总量的相当一部分 (22%)，用于此类激励系统。

新产品也将经历两个阶段。产品添加之初，协议不会允许针对该产品提出索赔申请。当产品只有少数发布者时，喂价不够稳健，因为每个发布者对聚合价格都有显著的影响。一旦产品有了足够数量的发布者，治理机制将通过投票，支持针对该产品提起索赔申请。但是，协议也将允许消费者为尚不支持索赔的产品支付数据费用，因为消费者支付数据费用的行为具有传递讯号的作用。例如，如果一个项目希望它的治理代币上线 Pyth，便可以利用这个机制向发布者提供额外的激励。

8 结论

本白皮书提出了一种实现准确、高保真的金融市场数据在链上的便捷访问和使用的预言机协议。协议旨在协调数据发布者和消费者，成为自我维持、去中心化的网络。面向消费者的数据质押机制是协议设计的一个关键部分，它将数据费用的一部分分配给发布者。Pyth 协议还旨在吸引提供高质量价格数据的发布者。协议的各项机制有助于防止恶意攻击者以各种方式操纵协议以谋取利益，例如操纵价格。